Fairness in the division and completion of collaborative work

Kelly, Ryan

Award date:
2014

Awarding institution:
University of Bath

FAIRNESS IN THE DIVISION AND COMPLETION OF COLLABORATIVE WORK

RYAN KELLY

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF BATH

DEPARTMENT OF COMPUTER SCIENCE

OCTOBER 2013

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my academic supervisor Professor Stephen J. Payne. This thesis would not exist without his guidance and ideas. He has been an excellent mentor and I am especially grateful for his enthusiasm and confidence in me. I must also extend thanks to Dr. Leon Watts, and to my undergraduate supervisor Dr. Jacqui Taylor, for encouraging me to follow this path. I would like to thank Professor Peter Johnson for feedback and encouragement during my PhD transfer meeting, and to my examiners Dr. Hilary Johnson and Professor Tom Ormerod for giving me new insights about old data.

For friendship and some much-appreciated research assistance, I would like to thank Laura Benton, Pawitra Chiravirakul, Daniel Gooch, Simon Jones, Andy Ridge, and Vicky Shipp. James Maidment also helped with some early analysis. Thanks are due to the rest of the HCI group at Bath, particularly Fabio Nemetz, Garry Ren, Em Srikulwong, and former member Dr. Geoffrey Duggan. I am also grateful to Martin Post, who provided some emergency technical support along the way, and to Andrew Chinery for (unknowingly) inspiring me to splash out on a new drum kit.

In February 2013 I was lucky enough to be accepted to the doctoral colloquium at CSCW, and I must therefore extend my thanks to panel members Geraldine Fitzpatrick, David McDonald, Pamela Hinds, Jacki O'Neill, Danyel Fisher, and Susan Wyche for their guidance and feedback. I also participated in the doctoral consortium at British HCI 2011 and I would like to thank Russell Beale and Liz Sillence for the experience. Thanks are also due to my peers, who provided stimulating conversation (and many laughs) at both meetings. Cheers to the SV crews at CSCW 2013 and 2014. Remember, I will always be the token Brit.

Last but not least, I would like to thank Caitlin Woods for lots of love and patience over the last three years. I am also very grateful to the EPSRC for research funding. My biggest thank you is reserved for my family. I would not be writing this if it were not for their everlasting support and tolerance.

Abstract

Fairness is an important concept that regulates many everyday transactions in human societies. While a large literature on the subject of fairness exists within the social sciences, the subject of fairness has not yet been explicitly addressed by researchers working in the area of computer-supported cooperative work (CSCW). One reason for this is that fairness may not seem especially critical for collaboration and, therefore, appears irrelevant to technological design. Yet perceptions about what is 'fair' often influence decisions concerning the assignment of tasks, the investment of effort, and the distribution of rewards during collaborative projects. Additionally, making accurate judgements about fairness can be difficult in computer-mediated settings where awareness of shared efforts is limited.

This thesis investigates the relationship between fairness and division of labour in collaborative work. We adopt a mixed methods approach to explore the prevalence of fairness in initial, distributive allocations of workload, as well as how fairness comes into play during the enactment of collaborative tasks. Our first study finds that initial allocations of work are made on the basis of fairness, and that perceptions about fairness in division of labour are linked to overall satisfaction with group outcomes. We then introduce a novel model of workload assignment based on the classic 'ultimatum game', and explore our model in a series of empirical studies using collaborative search tasks. Our findings provide further evidence of fairness norms in the allocation of work. We also report an emergent matching effect in participants' task completion times, which we take as further evidence of fairness in the enactment of collaborative workloads. We then report a qualitative study of collaborative search tool use in everyday tasks. This study draws attention to the importance of context in determining the extent to which people police fairness, and provides numerous implications for the design of collaborative search systems.

Our final thesis study explores the potential for supporting fairness during CSCW. We suggest that designers should provide awareness about individual contributions and should allow individuals to assess fairness via social comparisons. We report a study of how groups in the online game *World of Warcraft* use awareness mechanisms to keep track of individual contributions to the collective effort. Our study implies the potential utility of contribution meters while also shedding light on potential side-effects. The thesis ends by considering the implication of our results for current understandings of fairness. We propose a theoretical model that describes the process by which people assess fairness in the division and completion of collaborative workloads.

PUBLICATIONS

The contribution of the work presented in this thesis has been recognised through peer-reviewed publication in the following scholarly outlets:

- Kelly, R., and Payne, S. J. (2014). Collaborative web search in context: a study of tool use in everyday tasks. In *proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 807-819.

- Kelly, R., and Payne, S. J. (2013). Division of labour in collaborative information seeking: current approaches and future directions. In *proceedings of the 3rd International Workshop on Collaborative Information Seeking: Consolidating the Past, Creating the Future*, at CSCW 2013.

Summaries of this work were presented at the following doctoral consortia:

- Kelly, R. (2013). Fairness in the division and completion of collaborative work. In *proceedings of the 16th ACM Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 65–68.

- Kelly, R. (2011). An economic approach to studying division of workload in collaborative search tasks. In *proceedings of the 25th BCS Conference on Human-Computer Interaction (BCS-HCI '11)*. British Computer Society, Swinton, UK, UK, 539–542.

CONTENTS

List of Figures

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Thesis Overview

Working with others is a pervasive aspect of human life. Many endeavours are made easier by the involvement of two or more individuals, and some tasks are intrinsically collaborative to the extent that they simply cannot be completed alone (Roberts, 1991; Denning & Yaholkovsky, 2008). In the modern era, organisations leverage the power of groups to produce many complex outputs, and much of the work undertaken in pursuit of these outputs is mediated by computer systems. Yet developing systems to support collaboration is a significant research challenge, and part of that challenge involves understanding the nature of group work itself. For example, collaborative work has the potential be synergistic, in that workers can achieve outcomes that would not be possible alone (Schmidt & Bannon, 1992; Shah & González-Ibáñez, 2011). However, groups often face challenges that lead to shortfalls in performance, ranging from the inherent difficulty of managing distributed collaboration (Grinter *et al*., 1999; Gutwin *et al*., 2004) through to coping with social phenomena including shortfalls in collective effort (Latane *et al*., 1979) and conflicts between individual and collective goals (Symon *et al*., 1996). To compound matters further, technology itself can pose a barrier to collaboration, especially when there is little to no support for the behaviours that so naturally underpin work practices in face-to-face settings (Gutwin & Greenberg, 2002; Pinelle *et al*., 2003).

The subfield of human-computer interaction (HCI) known as computer-supported cooperative work (CSCW) seeks to alleviate some of the challenges that groups encounter in computer-mediated settings. While CSCW is very much concerned with the functional properties of groups and how technologies can help groups function (Ellis *et al*., 1991), the field also concerns itself with the design of tools that help people to work in harmony (Kraut, 2003) and further aims to examine the interplay between computer systems and organisational processes (e.g. Harper *et al*., 1991; Heath & Luff, 1992; Orlikowski, 1992; Heath *et al*., 1995). Striving toward these aims has required that researchers attend to group interaction and the processes that are integral to collaborative work more generally. For example, researchers have recognised that collaboration is a cooperative activity requiring awareness of one's colleagues (e.g. Gutwin

& Greenberg, 2002), division of labour among group members (Gaver, 1991; Mark *et al.*, 1996) and coordination of efforts to avoid conflict and prevent redundancy (Malone & Crowston, 1994; Bayerl & Lauche, 2008).

Although the need for CSCW systems to support constructs like awareness and division of labour is well recognised, collaboration is a complicated subject and there are many other processes that can influence the success of a group. One such issue, which has received very little attention from researchers in CSCW, is that of *fairness*. Fairness is an important concept that regulates many everyday transactions in human societies (cf. Binmore, 2010a). In the context of collaborative work, perceptions about what is 'fair' can influence individual decisions about how task components should be allocated; how much effort to invest in a common project; and, later, how rewards from group efforts should be distributed (Hertel *et al.*, 2002). While such issues are naturally quite difficult to assess, judgements of each might be especially difficult in computer-mediated environments where the visibility of one's colleagues is low and knowledge about efforts may be incomplete (Galegher & Kraut, 1990; Hertel *et al.*, 2003).

Within the wider social sciences, particularly psychology and economics, the subject of fairness has received considerable research attention. One reason for this is because human behaviour, in comforming to norms of sharing and fairness, fails to adhere to behavioural models that take selfish, egoistic greed as their most basic assumption. One such model is the neoclassical *homo economicus* model of economic man, where the canonical prediction is that people should behave selfishly and strive towards utility-maximisation above all else. However, hundreds of studies (see, e.g. Camerer, 2003, for a review) show that, during the various group tasks and social dilemmas used to investigate these models, people tend to prefer equity over selfish outcomes. In other words, people care about fairness and are content to punish those who do not behave in a suitably cooperative manner. These findings have led to multiple theories that attempt to reconcile fairness alongside other issues including reciprocity, competition, and rationality (e.g. Bolton, 1991; Rabin, 1993; Fehr & Schmidt, 1999; Levitt & List, 2007).

Despite this considerable body of work, researchers in CSCW have not yet addressed the subject of fairness in an explicit fashion. This may be because fairness does not seem like an especially important issue for technological design, yet a close reading of the CSCW literature reveals that fairness has arisen as a topic of concern in several studies of collaborative work (e.g. Galegher & Kraut, 1990; Dong & Fu, 2012; Merritt & McGee, 2012). In addition, problems related to fairness can also become salient during the design and evaluation of tools intended for collaboration (e.g. Morris & Horvitz, 2007). Other issues in group work speak directly to fairness; for example, the 'free-rider problem' occurs when "one or more members of a group do not do their fair share of work on a group project" (Brooks & Ammons, 2003, p. 1), a definition that in itself suggests non-compliance with fairness as the primary issue at hand. Taken together, these issues imply that fairness may be worthy of further study and could be at the heart of a potential design problem for CSCW systems.

This thesis explores fairness in collaborative work and CSCW by focusing on the division, and subsequent completion, of labour during collaborative projects. As we have seen, division of labour is an integral aspect of collaboration and groups need to make decisions about *who* will do *how much*, and of *what*, during a project. Such decisions may be taken in light of fairness preferences, and there may also be an expectation that the effort invested in a shared project will meet some standard of fairness. As mentioned above, striving for fairness may be especially difficult to manage in CSCW systems where awareness is limited. This thesis, then, has three key goals: 1) to identify the importance of fairness in collaboration through empirical studies of collaborative work, 2) to examine the extent to which the allocation and completion of workloads are guided by fairness, and 3) to explore how CSCW systems might be designed so as to allow groups to strive for fairness in the completion of their work. We address these subjects through a variety of empirical studies that use a combination of research techniques including questionnaires, interviews, and experiments.

The remainder of this chapter is structured as follows. Section 1.2 delineates topics that are, and are not, considered within the scope of this thesis. We then offer an overview of questions that the present work explores. These questions provide a backdrop to the topics that will be explored in later chapters. Section 1.4 then describes the methods employed in this work. Finally, Section 1.5 outlines the chapters that form this thesis, highlighting studies performed and the contributions made within each.

1.2 Research Focus and Scope

This thesis focuses on exploring fairness during the division, and subsequent completion, of work in collaborative tasks. These are two areas in which fairness norms may come into play during collaborative work. Given our focus, there are a number of aspects related to division of labour and fairness which will not be considered in this thesis. Some of these pertain to issues that we simply do not consider—others are more tenuous but may nevertheless be considered as potentially relevant to the issues of fairness and technology. Below we first consider what is meant by fairness within the context of this thesis.

1.2.1 Scoping Fairness

Fairness is an elusive term, and while most people can intuitively sense when they have, and have not, been treated fairly, many would struggle to articulate their understanding of fairness in precise terms. For the purposes of this thesis, our definition of fairness is as follows:

> *The quality of treating people in such a way that each receives his or her due in accordance with perceived entitlements, needs, and rights.*

The rationale for this definition is elaborated in Chapter 2, but in the meantime it is important to recognise that this definition is complemented by two issues. First, we argue that fairness involves a basic expectation of *equality* if there is no clear reason to behave otherwise. This appears to be true of Western cultures (Henrich *et al*., 2001), and means that, with all other things being equal, people expect to be treated the same, to be rewarded identically, and have equitable access to goods and resources (cf. Binmore, 2010b,a; Hobbs, 2010). The second issue is that equality is *not* the same as fairness. This is because true fairness is not about blindly following equality, but instead calls for consideration of other factors that would *legitimise* deviations from equality. As encapsulated within the definition, such factors might include an individual's entitlements, needs, or rights. For division of labour, this might mean that one individual has a legitimised *right* to do less work on a project, perhaps because he is in a position of authority over subordinates and must therefore focus on other tasks. In more egalitarian settings, however, there may be a basic expectation that each person contributes equally to the task—such a scenario is explored within this thesis through a study of student workgroups, as reported in Chapter 3. The corollary, then, is that the notionally 'fair' outcome in a given scenario is not necessarily the one that is most equal, but is instead the one that takes other issues into account.

In addition, it is important to note that it is explicitly *not* the aim of this thesis to state what is, and is not, fair. We take no moral standpoint on the subject of fairness because, in our view, fairness is in the eye of the beholder (cf. Wilkinson, 2008). Moreover, fairness is subjective and, as hinted above, can be influenced by cultural norms (e.g. Roth *et al.*, 1991; Henrich *et al.*, 2001; Buchan *et al.*, 2004; Oosterbeek *et al.*, 2004; Chuah *et al.*, 2007; Chen & Tang, 2009). This makes fairness all the more intriguing as a design problem for CSCW systems. In Chapter 4, we review a body of work from experimental economics that highlights the many ways in which fairness preferences are adjusted to legimitise unequal allocations of goods.

Further to these considerations, fairness is a nuanced term and can refer to different patterns of behaviour. We have stated that we are interested in the *division* and *completion* of workloads during collaboration. The former of these constructs aligns broadly with the topic of *distributional* fairness. The latter pertains to perceptions about whether the *contributions* made by each team member are perceived as fair. This differs to the typical construction of *procedural* fairness, which has more to do with whether or not an established set of procedures are fair and just (Lind *et al.*, 1990). An example is that of *due process*, referring to an established and consistent set of procedures that are followed in courts of law. Such procedures are in place to ensure that each individual is treated equally irrespective of the crime for which her or she may be on trial.

Additionally, this thesis shall not consider the topic of *retributive fairness*. Like other aspects of fairness, retribution is related to the concept of justice but pertains to how individuals should be punished for wrongdoings, i.e. whether or not a particular punishment is fair and whether an individual has been 'paid back in kind' (Miller & Vidmar, 1981). To use another example from law, it is commonly considered fair for the punishment to fit the crime; sentences should be commensurate with the severity of the act, the distress caused by the act, the extent to which the act was intentional, and perhaps also whether the pepetrator demonstrates remorse. In the context of collaboration, this could be relevant for the sanctioning of free-riding behaviours. While there is a large and ever-growing literature on social punishment of non-cooperators (see, e.g. Fehr & Gachter, 2000; Chaudhuri, 2011), retributive fairness has not arisen as a concern in the studies reported here and, therefore, will not be considered in this thesis.

1.2.1.1 Other Fairness-Related Issues in HCI

In addition to the issues outlined above, which primarily concern the method by which fairness is defined and the angle from which it is considered, there are some topics within HCI that could easily be regarded as relevant to the concept of 'fairness'. It is important to recognise that these will not be considered here. Examples include:

- Social issues over fair or equitable access to information technology, such as ICT for economic development and whether access to computer systems should more equitable irrespective of economic or social status. This might also be extended to the inclusion of underrepresented or minority groups in the development of technology, e.g. those with autism or other disorders (Benton *et al.*, 2012).
- The role of gender in IT, as pertaining to issues surrounding gender equality and the potential relevance of feminist theories to HCI (e.g. Bardzell, 2010).
- Ongoing debates regarding alleged tax evasion by large, multinational technology corporations (see, e.g. Commons Select Committee Report, 2013). This deliberate avoidance of tax might be regarded as unfair given the power of these corporations alongside their considerable influence on everyday life.
- The impact of work platforms and practices on employees, e.g. platforms for crowdsourcing and micro-taskwork like Amazon's Mechanical Turk.[1] Recent work has considered these platforms in terms of whether they are beneficial or detrimental to the professional development of remote workers, and whether or not the payoffs for both parties are constructed fairly (O'Neill & Martin, 2013).

Of course, this is not an exhaustive list, but the topics mentioned above are representative of the subject matter that might immediately come to mind when different members of the HCI community think about the term 'fairness'. This is again very much exemplary of the subjectivity associated with the word itself, and thus it is important to recognise that such topics are not considered within this thesis.

1.2.2 Scoping Division of Labour

For this thesis we define division of labour as "the act of breaking up a task for distribution among members of a group" (Foley & Smeaton, 2010). This definition does not entail that decisions have been made about *who* is most suitable for a given task, as might be the case during *task allocation*. Here we constrain our focus to the distribution of notional work units among members of a team. It is certainly true that the allocation of work in real-world settings will be based on task-related properties, and will account for issues such as skill, expertise, knowledge, and so forth. These issues are, in turn, very likely to influence what is perceived to be 'fair'. Additionally, allocation of work may be dictated by organisational policies,

[1] https://www.mturk.com/mturk/welcome

job roles, or status hierarchies (Bardram, 1997; Eason, 1996; Schmidt, 1994). However, the present studies cannot account for all of these issues in detail and thus we consider these as areas for future work. We restrict our focus to egalitarian settings that do not have predetermined methods for assigning work, as is the case with the sorts of student workgroups we study in Chapter 3.

1.2.3 Scoping Summary

In summary, this thesis is scoped to consider the following issues:

- Fairness as it relates to the initial division of work among members of a collaborative group. Hereafter we refer to this as *fairness in the division of work*.
- Fairness as it relates to the contributions made by members of a collaborative group during the enactment of divided work. Hereafter we refer to this as *fairness in the completion of work*.

1.3 Research Questions

The research presented in this thesis is guided by the following high-level research questions. These are not necessarily questions that we answer definitively; rather, we regard them as broad concerns that frame the studies reported herein. Our first question is: **RQ1: What is the relevance of fairness for collaboration and collaborative systems?** This question is explored broadly by all of the studies presented in this thesis. To provide a foundation for our efforts, understandings of what is meant by both collaborative work and fairness are developed in Chapter 2 of this thesis. Relevant literature is reviewed with the aim of drawing clear links between collaboration, fairness, and CSCW.

A second question we explore is: **RQ2: To what extent do people strive for fairness in the division of labour?** This question is explored through the empirical studies reported in Chapters 3, 4, and 5 of this thesis. Chapter 5 is especially relevant in this regard, where we present results that suggest fairness prevails in both the allocation and enactment of shared workloads. Additionally, a qualitative field study reported in Chapter 6 furthers this research question via a study of collaboration in everyday tasks.

A final question we consider is: **RQ3: How might collaborative systems be designed to account for fairness preferences in division of labour?** Our exploration of this question is based on two premises. First, that collaborative groups *do* desire fairness in the division and completion of their work, and second, that groups often find it difficult to ensure that work is completed fairly, due in part to a lack of awareness about the efforts and contributions of their collaborators. RQ3 is primarily explored in Chapter 7, where we aim to open a tentative design space for supporting fairness in CSCW systems.

In exploring these questions we also attend to various sub-concerns that arise as a result of our practical studies. These concerns expand the contributions of this thesis to other areas within CSCW. For example, the experiments reported in Chapter 5 use collaborative information seeking as an example task. By examining the strategies collaborators employ to prevent redundancy, we are able to suggest ways in which collaborative search tools could be designed to facilitate division of labour. Then, in Chapter 6, we retain the task of collaborative information seeking and explore real-world tool use as a major focus. This offers a range of tractable implications for the design of search systems.

1.4 Research Methods and Approach

This research uses a mixed-methods approach that draws on quantitative and qualitative techniques. Laboratory and field methods are employed, and several types of data are collected within each thesis study. The exact choice of methods is defined by the concerns of each study and the issues that each aims to explore. Our first study employs a survey to tap the opinions of students who recently engaged in collaborative projects. We use Likert-type scales to collect quantitative data and free-text responses to collect qualitative data. The approach of this study helps to ground this thesis in the concerns of real-world groups for whom fairness is a salient emotional concern.

The second approach used in this thesis (studies 2–5) is the framing of division of labour using an applied economic game. Such an approach is novel in the sense that it involves the application of game-theoretic concepts and methods to CSCW situations. Economic games have considerable currency in the social science literature; participants in such games are typically presented with decision scenarios that pit their individual payoffs against those of a group (Camerer, 2003). Individual decisions are more profitable in the short term but always have long-term negative consequences for the collective good, as is the case with the classic 'public goods game' where participants decide how much wealth to place into a common pool (Komorita & Parks, 1995). In our approach, we take one such game—the classic 'ultimatum game' (Güth *et al.*, 1982)—and apply it to the context of dividing labour. We use the model to capture distributional allocations of workloads made by participants engaged in a simple collaborative task. This approach offers a very clean exploration of fairness in the division of labour. And, like other economic games, it abstracts away from other potentially influential factors while allowing for legitimate between-study comparisons.

In exploring our economic game, we collect qualitative data that captures participants' motivations for dividing work and the strategies employed to avoid redundancy. Analysis of search protocols also provides quantitative data in the form of participants' task completion times.

Our third approach, as used in thesis study 6, comprises a field study that features real-world deployments of two tools designed to support collaborative information seeking. In contrast to our laboratory work, this study aims for a more naturalistic perspective on collaboration. Key methodological decisions in this regard were the recruitment of pairs of participants on the basis of existing information needs, alongside our choice to allow these participants to choose their own tasks and complete them within their own time. Thus, both information needs and the collaborative pairings themselves were pre-existing and intrinsically collaborative, providing high external validity. Our study findings are based on interview responses obtained from participants. This qualitative data is analysed using a grounded theory approach (Strauss & Corbin, 1998) and undirected, inductive coding (Elo & Kyngas, 2007). We chose this approach because we aimed to elicit a bottom-up understanding of collaborative search based on the dataset itself, rather than through confirmation of any pre-existing hypotheses. We had no particular preference for one outcome over another, and thus we adopted an approach that allowed the data to 'speak for itself' in favour of pre-existing classifications. Grounded theory was also beneficial in allowing us to maintain meaningful links between participants' work process and the way in which their assigned system was used to support collaboration.

Our final study uses several methods to explore how collaborative systems could allow users to make more accurate judgements about fairness. This is achieved through an examination of awareness mechanisms in an existing context; in this case, collaborative grouping in the online game *World of Warcraft* (WoW). Prior work identifies WoW as a context relevant to the core concerns of CSCW (Nardi & Harris, 2006; Bardzell *et al.*, 2008). We examine how players use *meters*, which visualise individual contributions to the collaborative task, as a means of gathering the necessary information to make fairness judgements. In terms of methods, the study is informed on three levels: first, the work draws on the author's own experiences of WoW—this approach is based on the premise that researchers should play the games they study (Williams *et al.*, 2006) and is an established procedure in studies of WoW (Bardzell *et al.*, 2008, 2012). Utilising the first-hand experiences of the researcher might be regarded as a variant of 'virtual ethnography' (Boellstorf, 2008), although it is important to note that the author's experiences did not enter into the dataset and merely served to inform the interpretation of said data. The study is further informed by participant observation, achieved through inspection of live, public webstreams of players engaged with current WoW content. This, combined with our earlier participation in the game world, provides the knowledge necessary to interpret our final level: interviews with actual players of WoW. Through our interviews we aimed to understand the derived experiences of players and the sociotechnical issues that may arise when making judgements about the contributions of team members. Our interview responses were again analysed via inductive coding, with codes shaped into high-level themes that allow for coherent interpretation of players' experiences.

In studies 6 and 7, all interviews were conducted, recorded, and transcribed by the author of this thesis. This procedure was adopted in favour of automated or paid transcription services because transcription is integral to the process of analysis; by coding the interview transcripts himself, the author was familiarised with the data at an early stage (Riessman, 1993) and was able to interpret meaning conveyed by participants' intonation and delivery (Braun & Clarke, 2006). This fed through into the transcription of interviews, which were transcribed so as to give a verbatim account of participants' statements for later interpretation, as recommended by relevant literature (Lapadat & Lindsay, 1999; Braun & Clarke, 2006).

1.4.1 Research Ethics

The work in this thesis has been completed in line with the University of Bath's Department of Computer Science 13-point ethics checklist. The checklist was reviewed prior to each study to check for potential violations. Appendix A provides an overview in this regard. While some of our procedures were relevant to items in the checklist, we did not identify any major problems, suggesting to us that our procedures and processes were ethically sound.

1.5 Thesis Outline and Contributions

This section provides a summary of the content and contributions made within each chapter of thesis.

Chapter 1 - Introduction

This chapter has introduced the problem domain and briefly discussed a number of topics that will be explored in later chapters. It has given rise to research topics that are explored within the thesis, and has outlined the methods employed during our empirical work.

Chapter 2 - Collaborative Work, Fairness, and Technology

This chapter aims to situate the contributions of the thesis within the context of computer-supported cooperative work (CSCW) while offering definitions of relevant terms. The chapter begins by introducing CSCW and reviewing its scope and primary research goals. We discuss taxonomies relevant to CSCW, and consider the difference between cooperative and collaborative work. Topics relevant to collaborative work are then introduced and discussed. Fairness is motivated as a relevant concern for collaborative work, particularly with regard to division of labour, and we review the concept of fairness at a broad (and necessarily selective) level. We suggest that perceptions about what is fair can impact both the division and and completion of collaborative tasks. Finally, the chapter is brought full circle by discussing CSCW literature where fairness is mentioned as a concern. The main purpose of the chapter is to provide a foundation for our empirical studies, and thus its primary contribution is an appraisal of topics relevant to work reported later in the thesis.

Chapter 3 - A Survey Study of Division of Labour and Fairness in Collaborating Teams

This chapter aims to identify the importance of fairness in collaboration via a study of students working on collaborative software development projects. Using a survey approach, thesis study 1 examines the link between fairness and overall satisfaction by asking students to rate how fairly they believe work was divided and completed within their groups. We observe a positive correlation between overall levels of satisfaction, both with processes and products, and perceived fairness in the distribution and completion of work tasks. The study also unveils some of the ways in which allocations can be adjusted from the point of equality without breaking fairness norms. Not only do the findings of this study raise fairness as a foremost concern in collaborative work, they also provide a very direct justification for the subject matter of this thesis.

Chapter 4 - Modelling Division of Labour as an Economic Game: Background

This chapter draws heavily on the economic literature to intoduce a novel empirical model for exploring fairness in division of labour. We begin by characterising division of labour as a process of negotiation, which we then propose to model by specifying an applied version of an economic game; in this case, the ultimatum game, a classic framing of negotiated exchange that has been used to study economic bargaining and fairness in a large number of studies. We present an in-depth review of the ultimatum game and its associated literature.

The contribution of this chapter is the distillation of our model alongside a review of the literature on ultimatum games. Our review offers insight concerning the effect of various methodological, structural, and demographic variables on fairness preferences. We also review prior use of economic games in HCI research, before considering how our model can extend prior work and enable future developments.

Chapter 5 - Empirical Studies Using the Division of Labour Ultimatum Game

This chapter presents four laboratory studies (thesis studies 2–5) that explore our novel DLUG paradigm. We draw on the review presented in Chapter 4 to inform our methodology and choices of independent variables. The studies all involve participants bargaining over the allocation of 10 items in a collaborative information seeking task, and we attempt to explore allocation behaviour through systematic manipulation of particular variables (e.g. knowledge of search topic). As this work is highly novel, our approach is incremental and begins with a very bare-bones scenario upon which we elaborate during each new study.

We report three main findings. First, the modal outcome of the work allocation is an even, notionally fair split. This occurs in all four studies, and persists even in a situation where the person allocating the work can behave entirely selfishly without the threat of punishment. Our second result concerns the completion of work: participants display a strong, yet entirely serendipitous, concordance in task completion times. That is, times are highly similar *within* pairs, but *between* pair times are often quite different. We regard this finding as further evidence of a desire for fairness, but this time in the overall completion of work. Finally, we examine the strategies used by participants to coordinate their work, finding that the use of coordination strategies depends on cognitive properties of the task at hand. We end the chapter by sketching some preliminary explanations for our results, and attempt to reconcile the findings of our studies alongside the extant literatures on the ultimatum game and fairness during social decision making.

Chapter 6 - Collaborative Search in Context: A Real-World Study

This chapter reports thesis study 6, a field study of division of labour and work completion during real-world collaborative information seeking tasks. In the study, two existing collaborative search tools are deployed to pairs of participants with pre-existing information needs. The study has dual contributions: first, we explore the potential importance of fairness during division of labour in everyday tasks. Second, by identifying how our chosen tools are used during collaborative search, as well as describing issues that embody particular design opportunities, we provide tractable implications for the design of collaboration information seeking systems.

Chapter 7 - Supporting Fairness Through Awareness: Contributions in Collaborative Gaming

This chapter brings together the results of our earlier studies to explore fairness as a design problem. Specifically, we suggest that designers can support fairness in two ways: by providing sufficiently detailed information about contributions, and by allowing individuals to engage in social comparisons over said contributions. We review approaches to facilitating awareness in collaborative tools (e.g. social translucence) and introduce the idea of contribution awareness. Then, in thesis study 7, we explore the use of two existing tools used to promote group awareness in the collaborative online game World of Warcraft. We report findings from interviews with players that offer two contributions: first, we detail how awareness about contributions supports participation and equality in Warcraft groups. Second, we describe a variety of sociotechnical consequences and corollaries that could impact the success of contribution awareness mechanisms, as applied to other CSCW contexts.

Chapter 8 - Conclusion

This chapter consolidates the findings of this thesis and offers a review of our research contributions. We integrate findings from across our studies to draw out three salient themes regarding fairness in collaboration. We then present a theoretical model that aims to account for the decisions an individual would take when making assessments of fairness. Finally, limitations of the thesis are identified and potential directions for future work are discussed.

The following chapter presents a general review of literature relevant to the concerns of the present thesis. The review focuses on three topics. First, we aim to understanding the nature of collaborative and cooperative work alongside the need to design effective technological support for collaboration. Second, we explore the meaning of fairness and its relevance to social interaction. Finally, we consider the link between fairness and CSCW, and the potential implications for collaborative systems design.

CHAPTER 2

COLLABORATIVE WORK, FAIRNESS, AND TECHNOLOGY: REVIEW AND SYNTHESIS

2.1 Chapter Overview

This chapter reviews literature relevant to the concerns of this thesis. The aim of the chapter is to provide the initial groundwork required to situate our research contributions within the agenda of computer-supported cooperative work (CSCW). Here we begin by focusing on CSCW and the nature of collaborative work—we introduce topics relevant to our exploration of fairness and division of labour, and consider appropriate definitions of terms. We then introduce the subject of fairness as a potentially relevant concern for CSCW, and review prior studies from the HCI literature that mention fairness as a topic of interest. In addition to what is presented here, relevant literature is reviewed in the background sections of other chapters: Chapter 4 offers an extensive literature review relevant to the ultimatum game, which then forms the basis of the empirical studies reported in Chapter 5. Literature relevant to the topic of collaborative information seeking is reviewed in Chapters 5 and 6, and the subject of 'awareness' in collaborative work is given an in-depth treatment in Chapter 7.

This chapter proceeds as follows. We first consider the scope of CSCW and review several taxonomies that allow us to introduce various phenomena relevant to collaboration. We articulate a basic distinction between collaborative and cooperative work, and introduce the concepts of coordination, awareness, common ground, and division of labour as relevant to this thesis. We then outline our position on the subject of fairness, before considering how fairness might be relevant to the subject of collaborative work. The chapter ends by examining relevant empirical studies and outlines a basic scheme for the rest of the thesis.

2.2 Computer-Supported Cooperative Work: Overview

The term computer-supported cooperative work, or 'CSCW' as it is now more commonly known, was first coined in a 1984 workshop co-organised by Irene Greif, then of MIT, and Paul Cashman, of Digital Equipment Corporation (Grudin, 1994b). CSCW initially focused on office automation systems and other 'groupware' platforms that were beginning to emerge at the time. Early definitions of CSCW reflected these developments (Wilson, 1991; Kling, 1993), but as the proliferation of computing technology gathered pace, so too did the need to understand the nature of group work itself. The early definition of CSCW offered by Bannon & Schmidt (1991) reflects this change:

> *CSCW should be conceived as an endeavor to understand the nature and characteristics of cooperative work with the objective of designing adequate computer-based technologies...The focus is to understand, so as to better support, cooperative work.* (Bannon & Schmidt, 1991, p. 3)

Bannon & Schmidt's emphasis on work, rather than technology, was progressive because designers were all too often focused on building systems simply because the technology to do so was becoming available (Grudin, 1994b). Such technologically deterministic approaches to development were suppressed by an era of studies focused on the social context of the workplace. Examples of early contributions include studies of air traffic control centres (Harper *et al*., 1991), public transport control rooms (Heath & Luff, 1992) and a securities trading floor in the City of London (Heath *et al*., 1995). Such studies were critical in not only recognising the interplay between individual and cooperative activities (Schmidt, 2000) but also served to highlight the nuanced relationship between work practices and collaborative technologies. For example, Orlikowski (1992) reported an ethnographic investigation of one company's adoption of *Lotus Notes*, an application intended for the curation of organisational knowledge in a shared database. Her study drew attention to conflicts between the assumptions of system designers and the organisational culture in which Lotus Notes was used. Work in the company under study was broadly cooperative but, because of an extremely high level of competition for promotions, there was little incentive to share information when promotion was awarded on the basis of individual expertise. Other studies revealed the potential for technologies to alter the outcomes of group work, both positively and negatively (e.g. Olson *et al*., 1992b).

The emphasis on understanding the nature of work has never left CSCW, but there has been a marked shift in terms of what is considered relevant to the discipline. This shift is perhaps reflective of technological proliferation more broadly. The early perception that CSCW was about "small, self-directed professional teams" (e.g. Kling, 1993) has rescinded in line with technological developments that can support collaborative work on a global scale (Grudin & Poltrock, 2012). Additionally, the definition of what is meant by 'work' has been widened to encapsulate a range of social behaviours that include, and are by no means limited to, volunteerism (Kittur, 2010), computer gaming (Brown & Baarkhuus, 2007), the life of the family (Kazakos *et al*., 2013), and even the business of maintaining romantic relationships (Scissors & Gergle, 2013). Nevertheless, while the scope of the field remains in flux, the focus on 'work' as the commodity of interest is still very much evident, and the fundamental research goals of understanding and supporting group work remain the same, at least at a very broad level.

While the breadth and interdisciplinary nature of CSCW is one of its main strengths, it is also its primary nadir in that it faces "the challenge of being multidisciplinary" (Grudin, 1994a). This partly relates to integrating a diverse array of perspectives into a coherent framework. This challenge is, however, not unique to CSCW and is true of HCI more broadly. A more pertinent bone of contention concerns the meaning of each of the terms within the CSCW acronym, and whether any of these terms (or indeed all of them) has lost its applicability in light of ongoing technological change (Grudin & Poltrock, 2012). In particular, the meaning of the second 'C' of CSCW has been a topic of considerable debate—should it be cooperative, collaborative, or something else? Grudin & Poltrock (2012) recount a 1988 panel session at the CSCW conference, where one individual argued that organisational behaviour is more complex than 'cooperation' and might be better described as 'conflictual' or 'coercive' work. While the latter remarks are somewhat facetious (though may perhaps be true of certain work situations) there does remain a need for precise terminology because the description of a given work situation "constrains the questions which might be raised and the answers which might be given about it" (Holand & Danielsen, 1991, p. 17). This chapter will not attempt to resolve the wider debate but will consider basic distinctions between cooperation and collaboration, a task that is important for the present thesis as we wish to consider the role of fairness in collaborative interaction. Furthermore, there is also a need to consider dimensions that influence group work more generally, e.g. whether work occurs in the same physical setting or is distributed across multiple locations. The following subsection reviews basic taxonomies relevant to CSCW. We then introduce various models of collaboration that allow us to introduce relevant terms and consider theoretical distinctions.

2.2.1 Taxonomies of CSCW Systems

Within CSCW, researchers have often found it necessary to reflect on the dimensions that influence the work of groups (Grudin & Poltrock, 2012). This has been achieved through the use of various taxonomies to characterise work situations and identify the sorts of technologies that might be used within a given setting. Perhaps the most basic of these taxonomies is the 'Four-Square Map of Groupware Options', which classifies CSCW systems with respect to time and space (Johansen *et al.*, 1991). The taxonomy is comprised of a 2x2 grid, with two characteristics used to classify CSCW systems:

Time. The form of interaction (synchronous versus asynchronous working). This concerns whether or not the contributions of participants are made in real-time, or at separate times.

Place. The geographical situation of the participants or users (distributed versus co-located working). This concerns the physical proximity of participants and whether or not each person is immediately accessible to others.

		Place		
		Same	*Different / Predictable*	*Different / Unpredictable*
Time	*Same*	Electronic meetings	Teleconferencing/Video	Multi-user seminars
	Different/ Predictable	Work shifts	Electronic mail	Forums
	Different/ Unpredictable	Team rooms	Collaborative writing	Workflow systems

Table 2.1: *Predictability in relation to time and place in CSCW systems, from Grudin (1994a).*

These distinctions are useful because the organisation of work across spatial and temporal boundaries influences the way in which groups organise their activities (Grudin & Poltrock, 2012). In particular, distributed collaboration is known to be difficult to manage due to issues including differences in timezones (Grinter *et al.*, 1999), a lack of visibility (or 'awareness') between colleagues (Gutwin & Greenberg, 2002) and increased feelings of social distance (Bradner & Mark, 2002). Such problems can impact the ability of groups to achieve productive outcomes; for example, groups may fail to leverage the capabilities of their team and unwittingly ignore distant collaborators (Bos *et al.*, 2006). Additionally, temporal and spatial boundaries have implications for the type of support designers might choose to include within a given CSCW system. Asynchronous systems, for example, might provide persistent records of collaborators' prior work sessions (e.g. Morris & Horvitz, 2007). Such features might, however, be less critical in systems intended for short-term, synchronous use.

The basic framework offered by Johansen *et al.* was developed further by Grudin (1994b), who sought to account for the predictability of collaboration with an extension shown in Table 2.1, which features canonical examples of CSCW systems (Grudin, 1994b). Predictable collaboration can be highly constrained, as when sending an email to a colleague and expecting it to be read within a day or so. On the other hand, tasks like collaborative writing are less constrained and can be open-ended; collaborators might know one another's locations but contributions might be expected within an unspecified timeframe. The predictability of work is important because it shapes the interactions that occur between colleagues and, in turn, might further influence the types of behaviours a CSCW system needs to support.

In general, these taxonomies offer researchers a convenient method of 'pigeonholing' CSCW systems (Grudin & Poltrock, 2012) and are useful for drawing attention to the arrangement of groups and their work practices with respect to time and space. But ongoing technological developments make these taxonomies less and less applicable. For example, it is not clear how these grids would account for mobile computing systems. Furthermore, as CSCW has developed, researchers have become increasingly concerned with understanding the nature of work itself rather than in categorising systems—this is because there is a need to distinguish between different work situations and identify whether behaviour should qualify as collaborative, cooperative, or indeed something else altogether.

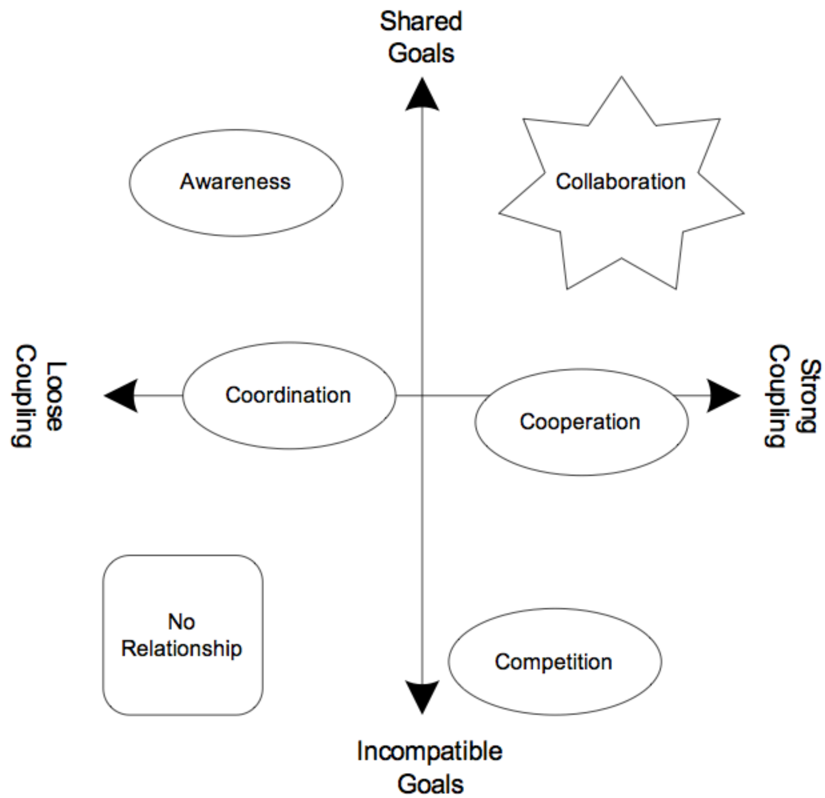

Figure 2.1: The ways in which group members interact, from Middup (2008, p. 39)

2.2.2 Models of Collaboration

As mentioned above, there is often a lack of consensus about the meaning of collaboration versus cooperation, as well as the extent to which they are independent or interrelated. The terms are often used interchangeably but, in our view, the two do not refer to the same behaviour. Here we consider some prior attempts at mapping out the conceptual relationships between the various elements of collaborative work. This allows us to segue into our own definitions, in turn providing deeper consideration of topics relevant to this thesis. It should be noted that we do not necessarily agree with the models we review here—their main benefit lies in introducing relevant terms and in identifying potential links among concepts.

Figure 2.1 displays a conceptual framework offered by Middup (2008). The model provides a diagrammatic overview of the various elements that Middup claims are involved in defining what is, and is not, involved in collaboration. Middup's definition of collaboration is "group work where the group members have some shared goal and the things that they need to do are strongly interdependent". The idea of a collaborative group as having a shared goal is one of the more consistent points of agreement within the literature (e.g. Symon *et al.*, 1996; Grudin & Poltrock, 2012) but the emphasis on collaboration as only occuring when subtasks are interdependent is peculiar. In general, the idea of interdependency refers to *coupling*, which pertains to "the degree to which people can work as individuals before needing

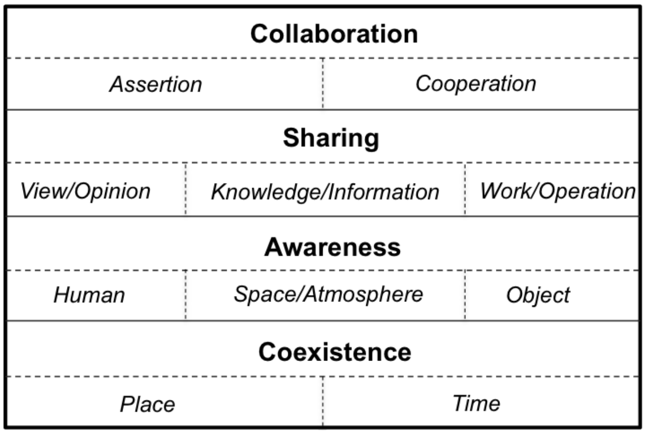

Figure 2.2: A hierarchical model of collaboration, from Okada (2007) (cited in Grudin & Poltrock, 2012).

to interact with another member of the group" (Pinelle *et al.*, 2003, p. 302). In tightly coupled work, group members must communicate frequently about various aspects of the shared project, whereas loosely coupled work "requires few interactions, and the communication that does exist is effortless, uncomplicated, and straightforward" (Neale, Carroll, & Rosson, 2004). In this way, Middup's definition precludes the classification of any loosely coupled work as collaboration, a position that is questionable because, as pointed out by (Olson & Olson, 2000, p. 162), collaborative projects are "not entirely tightly or loosely coupled. Various stages of the work are tightly coupled, and often there are stages where it is loosely coupled, where people... do the work in parallel". Collaboration, then, is not merely a 'state' but is also a process that is supported by division of labour—the act of decoupling tasks and distributing work among members of a group (Foley & Smeaton, 2010). Interdependency is simply a feature of particular projects, and we do not believe it is a prerequisite to collaboration. For example, searching and adding weblinks to a shared document during travel planning is very loosely coupled work, but participants might still regard themselves as collaborating because of their common goal.

A further problem with Middup's model is that he views coordination and cooperation as distinct from collaboration when, in fact, a majority of scholars regard the two as antecedent conditions that support effective collaboration. Figure 2.2 displays a good example in this regard, as outlined by Okada (2007) (cited in Grudin & Poltrock, 2012). Okada's framework offers a hierarchical model of collaboration, and is based on the idea that collaboration is strongly influenced by assertion and cooperation. Assertion in this case means exercising authority (perhaps akin to 'being competitive'), whereas cooperation can be taken as 'working harmoniously' with one's colleagues. Low levels of both "result in compromise; more assertion than cooperation results in collision; more cooperation than assertion results in concession; and high levels of both result in coordination" (Grudin & Poltrock, 2012, p. 19). Collaboration is then underpinned by communication, as occurs when members of a group share opinions, knowledge, or task-related information. The sharing layer is in turn underpinned by awareness of one's colleagues, of the environment in which work occurs, and of the various artefacts involved in the task. Awareness is then

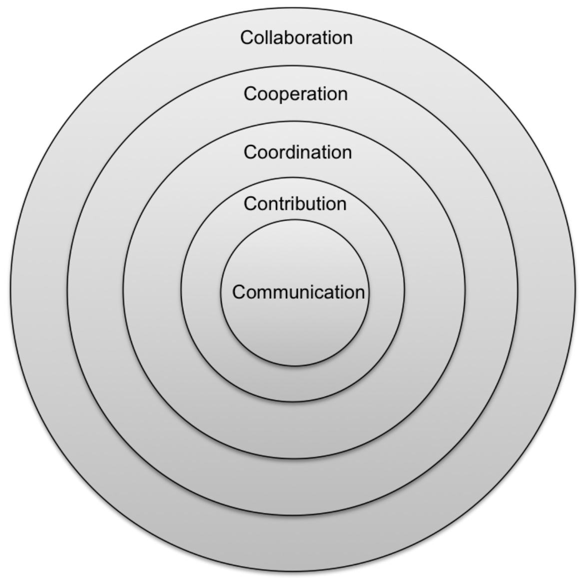

Figure 2.3: A set-based model of collaboration: "An inner set is essential to or supports an outer set", from (Shah, 2010, p. 6).

affected by temporal and spatial factors, as outlined in the earlier taxonomies by Johansen *et al*. (1991) and Grudin (1994a). (See subsection 2.2.1 above.)

The framework offered by Okada (2007) considers how group members interact but lacks any consideration of what collaborators actually *do* with the products of their individual work. Figure 2.3 displays a set-based model of collaboration from Shah (2010). The model has five components, each of which, Shah argues, supports the next. For example, coordination "is a subset of collaboration, which indicates that for a meaningful collaboration, there has to be coordination of people and events." (Shah, 2010, p. 6). The model proceeds from the innermost set to the outer, but one thing that is interesting about this model is that it incorporates the notion of contribution—the idea that individuals provide some meaningful input to a shared project, as opposed to simply 'working next to one another'. Shah also argues that progression from the innermost concept (communication) to collaboration involves progressively higher levels of *interaction, intent, trust, involvement*, and *awareness*. However, unlike the models described above, this framework does not account for the issue of a shared goal, a property that is regarded as essential for a collaborative group (Symon *et al*., 1996; Grudin & Poltrock, 2012).

The models we have reviewed here are certainly not the only characterisations of collaborative work, but by this point it should be clear that the literature harbours no consensus about how the various components of collaboration mesh together. There are, however, common notions about the behaviours it encapsulates and the processes that it typically involves. For example, Pinelle *et al*. (2003) characterise collaboration as involving a suite of low-level interaction primitives which, they claim, happen in co-located collaboration irrespective of organisational practices, personalities of group members, or even task type. They describe these behaviours as the 'mechanics of collaboration', which necessarily occur whenever two or more people work together. Table 2.2 presents a list of these behaviours. Examples include the transmission of messages, keeping track of what others are doing, negotiating access to shared tools or parts of the workspace, and exchanging or transferring objects and tools. These low-level actions then support higher-level phenomena including the development and maintenance of awareness, coordination of actions, communication, and division of labour. Pinelle *et al*. (2003) argue that a lack of support for these basic activities is what often makes CSCW systems difficult to use.

2.2.3 The Second C of CSCW: Collaboration or Cooperation?

Having considered a selection of prior models, here we attempt to offer a basic distinction between collaborative and cooperative work. Broadly, collaboration and cooperation can be taken to mean the act of 'working together' and, as evidenced above, it is certainly true that collaboration and cooperation share some common properties—most scholars identify that both involve communication and information sharing, for example.

For the purposes of this thesis, we define collaborative work as *"two or more individuals working together, making productive contributions to a common good in the presence of a shared goal"*. The need for a shared goal is one of the more consistent points of agreement within the literature, and we too ascribe to this notion. However, we would emphasise that the goal must be *shared* among team members, and knowledge that there is a shared goal must be mutual. In other words, collaborators have to be aware of the fact that they are collaborating, i.e. collaboration must be *explicit*. The shared goal itself can be more or less refined, but a group must operate under the knowledge that there is one. This establishes clear commonality of purpose (Bly *et al*., 1993) or what Tomasello (2009) refers to as *shared intentionality*.

Additionally, we believe that collaboration can be characterised by individuals making *contributions* to a common good, and we suggest that these contributions must be productive in the sense that they should be task-oriented and made in light of a group's shared goal. This makes collaboration distinct from situations in which people are merely working alongside one another, as might be the case in an office environment where people utilise shared artefacts but ultimately complete unrelated tasks in isolation (Rogers, 1993). We ascribe to the prior notions that collaboration is underpinned by *cooperation* and *coordination*. Cooperation in this sense constitutes agreement to behave in accordance with the behavioural norms or standards appropriate to the work situation. In other words, cooperation means "playing in the same game with others according to a set of behaviour rules" (Denning & Yaholkovsky, 2008). Refusal to behave cooperatively signals conflict, a mismatch in goals, and a general unwillingness to collaborate (Shah, 2010). Coordination is distinct from cooperation and refers to the job of "managing interdependencies between activities" (Malone & Crowston, 1990, p. 90). Successful coordination makes collaboration run smoothly by ensuring that contributions are appropriately timed, and that basic activities like conversation and access to shared resources occur without collision (cf. Pinelle *et al*., 2003). It is important to note that collaboration can still occur with poor coordination, but will be much less coherent and will likely break down quickly if coordination is not established.

Category	**Mechanic**	**Typical Actions**
Communication		
Explicit Communication	Spoken Messages	Conversational Verbal shadowing
	Written messages	Conversational Persistent
	Gestural Messages	Indicating Drawing Demonstrating
	Deictic references	Pointing & conversation
	Manifesting actions	Stylized actions
Information gathering	Basic awareness	Observing who is in the workspace Recognising what they are doing Identifying where they are working
	Feedthrough	Changes to objects Characteristic signs or sounds
	Consequential communication	Characteristic movement Body position and location Gaze direction
	Overhearing	Presence of talk Appreciating context
	Visual evidence	Normal actions
Coordination		
Shared access (to tools, objects, space, and time)	Obtain resources	Physically take objects or tools Occupy space
	Reserve resource	Move to closer proximity Notify others of intention
	Protect work	Monitor others' actions in area Notify others of protection
Transfer	Handoff object	Physically give/take object Verbally offer/accept object
	Deposit	Place object and notify

Table 2.2: The Mechanics of Collaboration, adapted from (Pinelle *et al.*, 2003, p. 288)

In contrast to collaborative work, *cooperative work* does not necessarily entail a shared goal of the sort that would be typical of collaboration. This is important because, when attempting to characterise a work situation, the idea of a shared work goal is not necessarily helpful. Often, organisational conflicts can exist which are incompatible with the assumption of a shared goal. For example, Symon *et al*. (1996) found that, in a hospital context, workers across departments had the same higher order goal (patient care) but this goal was often subsumed by other individual or group-derived goals that took precedence. In this way, personnel merely worked alongside one another and traded boundary objects when required during the process of patient care. Cooperation is thus a more appropriate term for such settings, in the sense that, while people are not in conflict, they do not regard themselves as 'collaborators' and do not necessarily pool contributions towards a shared goal.

Having considered a distinction between cooperative and collaborative work, we now look more closely at core aspects of collaboration that are relevant to this thesis.

2.3 Four Core Dimensions of Collaborative Work

This section delves into various properties of collaborative work. Our aim here to review some of the processes that are known to be involved in collaborative work and introduce the concept of division of labour as relevant to the present thesis.

2.3.1 Coordination

We noted above that coordination supports collaboration, but it is worth delving into coordination further to understand its integral role in work settings. Broadly, coordination refers to the management of activity within a group. Although various definitions have been proferred in the literature, the most commonly utilised definition of coordination is that offered by Malone & Crowston (1990), who define coordination as "the act of managing interdependencies between activities" (p. 90). Coordination therefore has a temporal focus in that it is about managing sequences of events such that those events occur seamlessly and without conflict. In work settings, this often relates to the need for certain jobs to be completed before others can proceed—managing and regulating these transitions is fundamentally what coordination involves. However, coordination is elaborated further by Rogers (1993) who states that the "interdependencies refer to common objects that are involved in some way in the activites" (p. 296). Common objects may refer to boundary objects such as medical records (Symon *et al.*, 1996) or collective cognitive artefacts like spreadsheets and paper documentation (Nardi & Miller, 1990). Such artefacts are often subject to simultaneity constraints, meaning that not everyone can make use of them at the same time. Access must therefore be coordinated, perhaps with individuals taking turns to access the shared resource. Such access is managed through 'coordinative behaviours', as outlined in the mechanics of collaboration (see Table 2.2).

As pointed out by Malone & Crowston (1990), we often notice coordination most when it is lacking; for example, when two objects collide, when two people bump into each other, or when two meetings are scheduled in conflict. Coordination, then, helps collaboration to proceed by avoiding conflicts. However, coordination is a cost *per se*. Bannon & Schmidt (1991) describe coordinative activities as a type of *articulation work*—that is, the 'work required to organise work'. Articulation work refers to the various activities in which collaborators must engage to ensure that their efforts add up to more than "discrete or conflicting bits of accomplished work" (Strauss *et al.*, 1995, p. 151). This requires additional effort, and can be detrimental to a group's progress if the overheads outweigh the benefits. Neale *et al.* (2004) state that the "overhead or operating costs involved in coordination is referred to as process loss, and distributed process loss is more costly" (p. 117). These overheads can stack up, requiring even further organisation that comes as a distraction from the business of actually getting the work done (Gaver, 1991). For example, Rogers (1993) studied software engineering teams and found that engineers spent a great deal of time engaged in *meta-coordination*—the organisation of various coordination protocols so as to avoid conflicts between each.

2.3.2 Common Ground

Common ground (cf. Clark & Brennan, 1991) refers to a body of knowledge that is shared among a group of collaborators. In order to have common ground, individuals must be aware that such knowledge is held in common by other members of their team (Olson & Olson, 2000). Individuals can make reference to their common ground during collaboration, and do so in the expectation that knowledge is shared. Common ground thus refers to the knowledge each person *believes* that others share in common with them (Olson & Olson, 2000; Neale *et al*., 2004). Common ground is continually updated during collaboration through a process called *grounding*. At the level of conversation, grounding refers to the "ongoing process of trying to determine whether what has been said has also been understood, comprising a joint effort on the part of everyone involved in a conversation" (Neale *et al*., 2004, p. 117). This process occurs collaboratively and iteratively (Clark & Brennan, 1991). Common ground is important because it is a precursor to effective communication and information sharing, which, as identified within the mechanics of collaboration (Pinelle *et al*., 2003), are foundational elements of collaborative work.

However, language is not the only means by which common ground is maintained. External representations also serve towards this purpose (Roschelle & Teasley, 1995) and common ground can be established according to a prescribed proforma or rubric, e.g. a task description that immediately gives collaborators some initial conception of a shared problem. Finally, testing and updating common ground can be expensive during distributed or computer-mediated work, meaning that common ground is generally easier to establish when teams are co-located (Olson & Olson, 2000). This is because common ground requires *awareness* of one's colleagues, and this is also true of many other collaborative behaviours.

2.3.3 Awareness

Early ethnographic studies of collaborative work revealed that, in situations where team members are physically co-located, people very naturally monitor and attend to the actions of their coworkers (Harper *et al*., 1991; Harper, 1992; Heath & Luff, 1992; Heath *et al*., 1995). This general practice of "taking heed of what is going on in the setting" (Schmidt, 2002, p. 285) is now recognised as a critical component of collaborative work, and has been described using the general moniker of 'awareness'.

Much like collaboration in general, awareness is a term that has often been used imprecisely due to a lack of basic consensus regarding what awareness is and what aspects of behaviour it encapsulates (Schmidt, 2002). As a term, awareness has its roots in the human factors literature, where the concept of 'situation awareness' has been especially influential (e.g. Endsley, 1995; Perla *et al*., 2000). Situation awareness refers to the "state of knowledge that an individual requires to operate or maintain a complex and dynamic system, such as an aircraft or nuclear generating station" (Gutwin *et al*., 1996a, p. 5). Situation awareness is especially relevant in military settings where personnel must remain aware of their team members, the nearby surroundings, and the presence of hostile threats. In CSCW, definitions of awareness tend to be more constrained than that of situation awareness. Perhaps the most widely cited definition is that offered by Dourish & Bellotti (1992), who describe awareness as *"an understanding of the activities of others*, which provides a *context for your own activity"* (Dourish & Bellotti, 1992, p. 107, emphasis as original). In essence, this definition tells us that awareness is about knowing what others are doing such that we can work out what to do ourselves. Context is used "to ensure that individual contributions are relevant to the group's activity as a whole, and to evaluate individual actions with respect to group goals and progress" (Dourish & Bellotti, 1992, p. 107).

Gutwin *et al*. (1996a) elaborate on this definition and identify four types of awareness relevant to CSCW: *Informal, Social, Group-structural*, and *Workspace* awareness. The first of these, informal awareness, relates to the general sense of who is nearby and what they are doing—"the kinds of things people know when they work together in the same office" (Gutwin *et al*., 1996a, p. 6). This type of awareness helps lubricate casual interactions and gives people a sense of presence in a social setting. The second, social awareness, refers to the information necessary to maintain person-person interaction—for instance, gestures, facial expression, and gaze, through to broader information like whether or not a person is interested in what we are saying and whether or not they are even paying attention. This type of awareness is important in any context where conversations or information-exchange is important. The third, group-structural, pertains to knowledge about one's team: who is who, what their roles are, what their responsibilities are, what are their opinions and stances on group issues, etc. This type of awareness helps collaborators determine appropriate actions that may depend on the intended audience (cf. Goffman, 1959).

Lastly, workspace awareness is a specific type of situation awareness (Gutwin & Greenberg, 2002), defined as:

> *"the up-to-the-moment understanding of another person's interaction with a shared workspace... [involving] knowledge about where others are working, what they are doing, and what they are going to do next."* (Gutwin & Greenberg, 2002, p. 412).

Gutwin *et al*. (1996a) regard this awareness as critical due to the situated nature of collaboration in a workspace—"when interaction happens in a workspace, maintaining knowledge about others' interaction with the space and its artifacts becomes highly relevant" (Gutwin *et al*., 1996a, p. 6). In other words, workspace awareness is not solely about understanding what is in the workspace but instead pertains to what people are actually *doing* with the things that are present within it. For our purposes, workspace awareness is useful because it accounts for general aspects of awareness, like 'who is who' and 'who is around', while stressing the importance of what the group is doing, how artefacts are being manipulated, and how the task is progressing within the context of that workspace (Gutwin & Greenberg, 2002).

It is also important to note that awareness is both a process and a product. As Gutwin *et al*. (1996a) describe, "the product is the state of understanding about another person's interaction with the workspace, that allows people to interpret events, anticipate needs, and interact appropriately. The process is the continuous cycle of extracting information from the environment, integrating this information with existing knowledge, and using that knowledge to direct further perception" (Gutwin *et al*., 1996a, p. 4).

The subject of awareness is given further treatment in Chapter 7 of this thesis, but for the time being it is important to recognise that awareness is relevant to a wide variety of behaviours in collaborative work. These range from basic features of coordination that require knowledge about who is involved in the collaboration, what their role is, and so on, through to cooperative acts such as giving assistance or offering clarifications (Gutwin & Greenberg, 2002). Such behaviours are enabled by *information gathering*, as outlined by the mechanics of collaboration related to communication in Table 2.2.

2.3.4 Division of Labour

A final aspect of collaboration we consider here is division of labour, a subject that is a primary focus for the studies described in this thesis. Here, we use division of labour to refer to "the distribution of work around members a group" (Foley & Smeaton, 2010). A similar term is 'task assignment', which might be regarded as a more specific type of division of labour focused on 'finding the right man for the job' (cf. Malone & Crowston, 1994; Labella *et al*., 2006). We prefer the term division of labour as we do not wish to assume that there is an appropriate individual for a given task, nor that tasks could even be assigned in this way—some collaborative projects may simply consist of notionally similar work items that can be completed by any particular individual within a team. Additionally, division of labour does not only refer to the distribution of explicit task labour but can also be regarded as relevant to the distribution of *cognitive* labour. As identified by (Olson *et al*., 1992a), dividing up the cognitive work means that each individual "does not have to cycle through all the criteria to ensure a good decision but, rather, can depend on others to share the burden... and release their cognitive energy to attend to other things" (p. 369).

In terms of its role in collaboration, previous work has recognised the variety of benefits conferred by division of labour. For example, tasks can be broken down into independent chunks, which may then be distributed across a team and completed in parallel (Mark *et al*., 1996). This parallelisation can overcome problems that sometimes hinder group work, by, for example, alleviating the cost incurred when groups working collectively must wait for their slowest member (Mark *et al*., 1996). Planned divisions of labour also serve as a status overview, bringing structure and coherence to individual activities (Bardram, 1997).

However, division of labour can be difficult to establish, and this is especially true of more complex projects (Gaver, 1991). A particular problem may be that the components of projects are tightly coupled. We noted earlier (see subsection 2.2.2) that coupling refers to the interdependencies among subtasks. In general, the state of divided labour is characterised by loose coupling, but the initial separation of subtasks can be problematic because some components may require completion before others can be started, or particular components may be interrelated to the extent that they are difficult to tease apart. Large software development projects are one good example; such projects typically involve many hidden dependencies that can make the work of organising division of labour an arduous task (Herbsleb & Grinter, 1999). In such cases, collaborators need to invest greater effort to not only separate the project into manageable chunks, but may also require greater effort to build awareness, maintain coordination and, eventually, integrate individual contributions into a coherent whole. Division of labour can be made additionally difficult by distance work (Grinter, 1996; Grinter *et al*., 1999), with groups needing to establish explicit coordination protocols (e.g. turn taking) in order to avoid redundancy (Mark *et al*., 1996; Olson & Olson, 2000; Schmidt, 2011). Protracted management of collaborative activity can also be a cognitive load ('collaborative load') (Fidel *et al*., 2004).

Detailing these concerns allows us to characterise division of labour as, very broadly, involving three phases, each of which pertains to the general activity of project completion (cf. Kraut *et al*., 1990; McGrath, 1990; Galegher & Kraut, 1994; Schmidt, 2011). These are as follows:

Planning, which is related to project initiation in that it requires settling on a shared interpretation of the problem (Galegher & Kraut, 1994) but also involves the initial distribution of work among members of a group. Collaborators may need to agree who will do how much of what, and perhaps also how tasks will be done (Freidson, 1976; Strauss, 1985). This phase may require frequent communication (Gaver, 1991) and careful decoupling of subcomponents, depending on the extent to which subtasks are interdependent.

Enactment, referring to execution of the work, where coworkers may transition between periods of focused collaboration to brief interactions requiring only minimal awareness (Gaver, 1991). The latter circumstances involve completion of work individually. However, tasks may be reorganised on an ad-hoc basis in accordance with project progression or situated action (Suchman, 1987).

Integration, where, as tasks are completed and a project nears its end, collaborators may need to weave their individual components into a coherent whole (Galegher & Kraut, 1994). The extent to which reintegration is necessary will depend on the initial coupling of subtasks; some collaborative efforts can simply be aggregated with minimal effort, whereas others may require careful integration.

The temporal organisation of these stages is intended as a general picture of how division of labour might occur, and thus it is not the only way in which work can be organised. Collaborators might, for example, be able to contribute bits and pieces of work as the project progresses, rather than delaying contributions until the very end of the project. The initial division of tasks might also be influenced in various ways, especially by the issue of *control*. Assignment of work can, for example, be guided by organisational protocols, power differentials and status hierarchies (Eason, 1996; Bardram, 1997); perceptions about gender roles (Mikula, 1998); job roles that prescribe assignment of specific subtasks (Symon *et al*., 1996; Poltrock *et al*., 2003); skills, knowledge, and individual expertise (Grinter *et al*., 1999); or standard operating procedures (Grinter, 1996).

2.3.4.1 Division of Labour and CSCW

Division of labour was very quickly identified as a problematic issue for CSCW systems, in the sense that technology can provide support for divided work but must also be sensitive to the need for fluid organisation of tasks (cf. Suchman, 1987). Two early examples highlight these issues. First, Rodden (1991) found that a lack of task partitioning was a problem in early collaborative hypertext editing tools. He concluded that there should be "some mechanism that allows authors to work together in independent hypertext partitions without risk of interference, and then to allow those independent partitions to be joined at carefully controlled intervals" (Rodden, 1991, p. 21). Second, systems that assign workloads rigidly can be problematic. Eason (1996) identified problems with a mobile data link used to manage the assignment of emergency and non-emergency tasks to electrical engineers. Rather than asking a foreman to delegate roles, as had been standard procedure for many years, the new system used call centre clerks as intermediaries to send tasks directly to engineers. This was controversial because tasks could only be 'delegated' and engineers could not respond to task assignments by accepting or rejecting work assignments. This caused them to lose any discretion over the acceptance of jobs and their own work schedules, in turn leading to conflicts when engineers could not meet the scheduled work assignments.

While distribution of control can determine the assignment of work in some settings, many groups are initially unstructured in their division of labour. Olson *et al*. (1992a), for example, studied design teams and identified that such teams spend about 10% of their time coordinating and organising labour. In such cases, which may often involve relatively egalitarian power structures, groups must discuss the assignment of work and achieve consensus about who is responsible for which aspects of the project. In line with the concerns of the present thesis, it is reasonable to expect that fairness could play a role in the allocation of tasks (Hertel *et al*., 2002) and also the way in which work is completed. The exact influence of fairness is, however, currently unknown. In the following section we consider the meaning of fairness and its potential relation to division of labour.

2.4 On Fairness: A Short Consideration

This section will consider the basic meaning of the term 'fairness'. In particular, we shall make an effort to distinguish fairness from related concepts like *equity* and *equality*. This is done so as to bring precision and consistency to our use of these various terms throughout the thesis. Before proceeding, however, we must draw attention to several caveats. First, we shall not review everything that can possibly be said about fairness; such an endeavour is not necessary for our purposes and would likely take a thesis of its own to achieve. We recognise that large bodies of work exist on fairness within psychology, philosophy, and economics—Chapter 4 of this thesis provides a closer look at the latter of these literatures.

Second, it should be noted that it is explicitly not the aim of this section, nor this thesis in general, to discuss what is, and is not, fair. We are agnostic with regard to fairness and adopt the view that fairness "is in the eye of the beholder" (cf. Wilkinson, 2008, p. 330). In other words, we do not seek to take any particular moral or philosophical standpoint on the issue of what is 'right' or 'wrong' when determining fairness. Instead, we shall aim to acquire a sufficient understanding of how people generally think about fairness, why it can be problematic to define, and how fairness is important in a general sense.

Because fairness is an extremely broad topic, we narrow the present discussion by considering fairness as it applies to the subject matter of the present thesis, i.e. division of labour during collaborative work. This topic can be roughly equated to the subject of *distributive fairness*, which refers to the process by which resources are distributed among group members (Deutsch, 1975; Hertel *et al*., 2002). We align this aspect of fairness with initial divisions of labour, where groups must decide *who* should do *how much* of *what*. Such decisions may be taken in line with what is perceived as 'fair'. Our second topic of concern is fairness during the completion of work. In this case, fairness may arise if certain team members fail to conform to the standard set within the initial division of labour, as occurs with the well-known free-rider problem (Brooks & Ammons, 2003). Alternatively, individuals may come to realise that workloads are unreasonably disproportionate, unsuited to the skills of assigned workers, or require reworking to become satisfactorily fair. There may also be some general desire for fairness in terms of *how* the work is enacted; for example, with reference to time on task, or the amount of effort invested into the shared project (Jackson & Harkins, 1985).

2.4.1 Defining Fairness

Fairness is one of the most important values in human society (Wierzbicka, 2006; Binmore, 2010a). Most of us are very well-acquainted with the idea of fairness; it is very common to hear the exclamation "that's not fair!", especially if small children are around at the time. Indeed, one need not look far to find mention of fairness in everyday life: from the *Fair Trade* movement to *fair use* policies, to institutions like the Office of *Fair Trading* and the *Fair Play* leagues in English football, it seems clear that fairness is pervasive. Yet understanding what is "fair" is a complex problem; fairness itself is an elusive term, with a subjective nature that can be difficult to translate between languages (Wierzbicka, 2006). And while most people can intuitively sense when they have, and have not, been treated fairly, many would struggle to articulate their understanding of fairness in precise terms. Given that we have offered fairness as a topic of concern for this thesis, it is crucial for us to consider the meaning of the word if we are to offer consistent discussion and interpretation within later chapters.

At a very basic level, fairness tends to have some degree of association with the idea of equality, and is often used interchangeably with terms like equity and egalitarianism. Indeed, a simple Google search for a synonym of fairness provides the word 'equitable' as one answer, with a further search for a definition returning *"treating people equally without favouritism or discrimination"*.[2] A simple definition of fairness, then, might be that what is fair is simply 'what is equal'. This represents perhaps the most straightforward fairness rule, and, for the most part, tends to be satisfactory in everyday situations, so much so that people are known to rely on a 'general equality algorithm' (Hertel *et al*., 2002) that prescribes equal allocation of

[2] Source: Oxford Dictionary, http://oxforddictionaries.com/definition/english/fair?q=fairness#fair

resources to individuals in exchange situations (Allison & Messick, 1990). Yet while most people would probably agree that fairness does have at least some basis in equality, it is not necessarily true that equal outcomes are always satisfactory. This is because equality, in the strictest sense, is inflexible and leaves no room for adjustment—equal is equal, and that is that. What is regarded as *fair*, on the other hand, is often open to interpretation, and may not necessarily be synonymous with equality all of the time.

If fairness is not merely about equality, then what is it about? To answer this question, it useful to think of fairness in terms of its role as a social norm. A norm can be defined as behavioural standard that creates "shared expectations about how all group members ought to behave" (Levine & Moreland, 1998, p. 427). In line with Binmore (2010a), we begin by regarding equality as the default position for the fairness norm. This is because, in line with our statements above, equality is a status quo that is acceptable from a moral point of view (Rawls, 1999) and is regarded as the 'fairest' choice in situations where there is no clear reason to behave otherwise (Allison & Messick, 1990). This is at least true of Westernised cultures (Henrich *et al*., 2001; Wierzbicka, 2006), and means that, with all other things being equal, people expect to be treated the same, to be rewarded identically, and have equal access to goods and resources (Hobbs, 2010).

From the basic position of equality, fairness involves taking into account issues that affect perceptions about what each person is perceived to be *due* from a given situation. Fairness, then, is about *proportionality of treatment*, in the sense that it requires one to attend to relevant factors that might provide rationale for deviations from equality (cf. Hertel *et al*., 2002). If said deviations are judged as *acceptable* by the involved parties, then the outcome is likely to be perceived as fair. Although we cannot provide an exhaustive list, such factors include entitlements, needs, or perceived rights to a given resource. For example, it is generally considered acceptable for those in the most dire of medical straits to have preferential access to emergency healthcare. In this case, certain individuals have an established *need* over others, in turn suggesting that it is *fairer* for them to receive preferential treatment over those with lesser needs. Similarly, perceived *entitlements* might influence decisions about what is fair when goods are allocated. For example, if someone has contributed more work to a collaborative project, he or she might perceive a right to greater credit as a truer reflection of contribution. Factors more relevant to the context of division of labour might include differential skills or abilities that guide the assignment of tasks to particular individuals. In an egalitarian setting, there may be no clear reason to stray from equality in the assignment of duties, but status hierarchies or positions of authority might legitimise the ability of one individual to 'dictate' tasks to another, as would occur with a manager and his subordinates.

To provide a precise definition, then, we suggest that fairness relates to:

> *The quality of treating people in such a way that each receives his or her due in accordance with perceived entitlements, rights, or needs.*

This definition allows us to begin with the basic jumping-off point of equality because, in line with societal norms, people generally perceive themselves as 'entitled' to certain considerations, most often in the form of equal treatment when there is no reason to behave otherwise (Allison & Messick, 1990).

From there, further factors may shift perceptions of fairness in one direction or another. As outlined above, specialist efforts, needs, or mitigating circumstances will affect perceptions about what is due to each person. For the context of division of labour, assignments of work can be influenced by a variety of factors, and these factors are roughly equivalent to the concepts given in our definition above, in the sense that a status hierarchy may imply entitlement to dictate work, whereas imbalance in expertise might imply that an individual needs to handle (or perhaps *should* handle) more of one task than another.

What is helpful about our definition is that it allows us to specify what is meant by fairness while, at the same, offering an opportunity to consider what is meant by unfairness. Simply put, any outcome that disregards the perceived rights or entitlements of an individual is likely to be judged as unfair. Broadly, fairness can be linked to the concept of justice (cf. Rawls, 1999) in that what is fair is often perceived as the 'just' or 'right' course of action with respect to laws or particular moral codes. The factors mentioned above may serve to *legitimise* deviations from equality without invoking moral indignation. On the other hand, deviations from equality where such factors are ignored or do not exist are likely to be perceived as *illegitimate* and, hence, highly unfair. Specific illegitimate factors might include pure selfishness or, in collaborative work, laziness or reluctance to participate in the task. Such concerns are at the heart of the free-rider problem, which "occurs when one or more members of a group do not do their fair share of work on a group project" (Brooks & Ammons, 2003), a definition that in itself taps failure to comply with fairness norms as the primary issue at hand.

However, what makes fairness so slippery is that it is always subjective, and this is very much reflected within our own definition. There is no true way of specifying exactly what is 'fair', and this is one of the reasons why forming agreements about what is fair can be problematic—individuals may place greater weight on some issues over others, and the claims of one individual about entitlement may not align with those of another. Similarly, the subjective nature of fairness is exemplified by cultural variation in fairness norms (e.g. Henrich *et al.*, 2001; Chuah *et al.*, 2007) and the fact that perceptions about fairness can change over time in line with wider shifts in societal norms. For example, up until the 19th century it was considered fair for people to be enslaved or lynched on the basis of skin colour. Such behaviour is now illegal and abhorred in modern society. The point here is that perceptions about what is 'due' are malleable and can differ from person to person. This is very much evident in modern life as well: for example, it was, until recently, considered acceptable for chief executives of large companies to be remunerated with very generous bonuses, even if those bonuses were awarded on the basis of relatively poor performance or even failure. Public perception of such disproportionate reward systems is changing, especially in light of recent economic woes attributed to individuals in positions of power within the financial sector.[3]

In defining fairness, we have touched on the issue of equality as our basic premise. It is worth noting that both fairness and equality are highly related to the principle of *equity*, which refers to the allocation rule of ensuring that outcomes are proportional to the contributions of group members (Levine & Moreland, 1998). We touched on this rule in our discussion above, and although our definition of fairness encapsulates equity, it is worth noting that equity might be regarded as insufficient compared to fairness because there

[3]See, e.g., http://www.bbc.co.uk/news/uk-16783571

may be circumstances that prevent individuals from contributing as they would like. For example, a colleague could have a physical disability that limits the speed at which he or she can complete work—one might regard it as unfair for this person to be penalised with lower remuneration, especially if he or she has made a best effort to contribute to the task, because the problem is beyond his or her control. Thus equity is often insufficient as an allocation rule, perhaps making fairness the better choice.

One final issue worthy of note is that fairness is fundamentally social, in that people would not care about fairness if they were not interested in the circumstances of others (Binmore, 2010a). This is true in terms of comparing one's own outcomes relative to those of specific referent others, as well as to how individuals are treated in society as a whole. Thus, *social comparison* (Festinger, 1954; Knez & Camerer, 1995; Cohn *et al.*, 2009) is an important driver for fairness—fairness judgements are always made relative to some standard, and that standard is often the situation enjoyed by others. If we see that the ratio of our own inputs and outputs is equal to the ratio of inputs and outputs of others (cf. van den Bos *et al.*, 1997), we are likely to judge an outcome as fair provided that it also accounts for other relevant factors. But if we see preferential treatment afforded to others for no clear reason, we might be inclined to denounce such treatment as very unfair. Within this thesis, the idea of social comparison is important because it is raised repeatedly in later chapters, and forms the basis for our own arguments about how fairness judgements might be supported in collaborative systems.

2.4.2 The Value of Fairness: A Basis for Cooperative Interaction

We have noted that, in everyday society, fairness is manifested in the form of broader social norms that govern how we interact with one another and the institutions that surround us (Binmore, 2010a,b). Many everyday behaviours can be construed as conforming with fairness norms: for example, the simple act of waiting in line can be viewed as conformity with fairness because, by joining a queue and standing patiently behind the person in front, we acknowledge that the person has the right to be served first, established on the basis of the fact that he or she joined the queue before we did. To ignore that right would be 'unfair', and although there is nothing stopping us from non-compliance with fairness norms, what usually puts us off is the threat of *reciprocity*, the general maxim of 'doing unto others as they do to us' (Fehr & Schmidt, 1999). Here, reciprocity may come in the form of punishment, perhaps in the form of an angry comment, ejection from the queue, or even violence that could lead to physical harm.

The idea of reciprocity helps us explain why people choose to conform to fairness norms. By treating others as we would hope to be treated, we in turn increase our own chances of receiving similar treatment. By recognising the claims, entitlements, and needs of others, we in turn can draw on similar rationale when it is our own turn to make use of a resource. And by recognising non-conformity to said norms, we punish others so as to make them comply with our expectations about fairness and what is right in our societies. Fairness is thus intrinsically related to cooperation. Cooperative behaviours keep our society ticking over in an amicable fashion—social order would not last long if conflict arose every time one encountered the sorts of picayune problems that arise on a day-to-day basis. For collaboration, fairness might be relevant to the recognised prerequisite of cooperative behaviour. It almost certainly links to the idea of 'going along with established procedures' (cf. our earlier discussion of cooperation) because, by playing fair, one is conforming to established norms of interaction. In the following section, we consider prior studies within CSCW where fairness has been mentioned explicitly. This leads to a consideration of our own research aims.

2.5 Fairness and CSCW

Despite the importance of fairness for our everyday social interactions, there appears to have been little considered exploration of fairness within CSCW and HCI, meaning that there is no immediately identifiable body of work from which we can draw. One possible reason for this is that fairness, beyond its potential role in division of labour, may not initially seem like an important issue for technological design. Yet a close reading of the literature reveals that fairness has been raised as a concern in a small number of studies—it is simply that no work has yet attempted to unite these disparate works into a coherent whole. This is perhaps because fairness, in the limited cases where it has been mentioned, tends to appear well down the list of concerns for the studies in question. In other cases, papers sometimes address the subject of fairness implicitly or obliquely. Here we offer a short appraisal of studies that touch on the subject of fairness, as identified by the author of this thesis. We begin with work that mentions fairness explicitly. We then consider other issues where fairness might be putatively relevant for CSCW.

2.5.1 Studies Mentioning Fairness

An early study by Galegher & Kraut (1990) (see also Galegher & Kraut, 1994) reports an experiment on group writing. Students completed a group assignment over a period of two weeks in one of three conditions: sole reliance on face-to-face communication, sole reliance on computer-mediated communication (text chat), or computer-mediated communication complemented by telephone conversations. The researchers aimed to gauge the effects of these communication channels on nine dependent variables, one of which was the perceived fairness in each group's division of labour. (The fairness measure was not, in this case, the primary concern of the research). Galegher & Kraut found that groups working face-to-face perceived greater fairness in their work process than those in the two computer-mediated groups. This effect was particularly pronounced while groups were engaged in more execution-oriented stages of work, as opposed to initial planning. For the present thesis, these findings are interesting as they suggest that groups may find it more difficult to make judgements about fairness when the completion of contributions is obscured by technology.

Fairness is mentioned in early CSCW literature on information sharing and privacy. For example, Kling (1993) discusses the potential implications of increased data capture and its potential implications for CSCW in terms of how information can be used ethically and safely. Similar concerns were expressed by Harper (1996) in a study of active badges—small software systems that remotely monitor the whereabouts of colleagues. Such badges were valued highly by those whose job it was to keep tabs on people (e.g. secretaries) but were regarded with suspicion by academics and research staff due to the perceived unfairness of increased scrutiny.

To find other studies that mention fairness as a concern, one must look much closer to the present day. Dong & Fu (2012) mentioned fairness in a study of video-mediated communication; they argue that video is better for conflict resolution because it promotes turn taking and equity in participation, in turn allowing for 'fairness' in negotiation. They conducted a study that required pairs of participants to achieve consensus in an appointment-scheduling task using either text-, audio-, or video-based communication. The study revealed that participants were more likely to reach a consensus when information exchange was managed in smaller chunks, and that this was best achieved via video. Since a more balanced exchange led to equality in the negotiation, Dong & Fu suggest that CMC tools might somehow be designed to foster a feeling of fairness. It should be noted, however, that the authors do not provide any clear design guidelines on this matter.

Fairness concerns also arose in a recent a study by Merritt & McGee (2012). They sought to explore how behaviour differs in cooperative computer games when people interact with human or artificial teammates. Participants played a simple game where they could perform a "yell" to protect their teammate from enemy gunfire. Each participant played in two different sessions: one with an artificially intelligent (AI) partner, and another with a human player who was, in reality, also an AI agent, but participants were not aware of this during the experiment. Participants self-reported a higher instance of yelling when playing alongside a 'human', yet log data showed that protective events were actually *lower* in this condition. In other words, participants protected the AI teammate more often but were not consciously aware of this behaviour. Merritt & McGee interpret this result in terms of fairness, suggesting that interaction with the human player leads to concerns about equality and a desire to 'match the behaviour' of one's counterpart. Such concerns were apparently less relevant when players were not concerned about being evaluated by an artificial teammate.

One might notice that there appears to be little consistency in the specific subject matter of the studies described above. The one theme that does run common, however, is that each is broadly concerned with interaction between two or more individuals, as mediated by computer technology. The fact that each study addresses a different aspect of HCI suggests to us that fairness issues may be more pervasive than first thought—different aspects of fairness may be important in a wide variety of interactions.

2.5.2 Other Related Work

In addition to studies where fairness has been mentioned explicitly, there are some other topics within HCI that touch on issues related to fairness. Of course, there are many areas that could be tenuously linked to the subjects one typically considers when talking about fairness. We do not review such work here—instead we briefly consider work that has mentioned the term fairness but not 'CSCW'.

Within some areas of HCI literature, there appears to be an implicit assumption that equitable participation is somehow preferable when groups work together through technology (e.g. Rogers *et al*., 2008; Fleck *et al*., 2009). Such arguments are predicated on the difficulty of collaboration when technological resources are limited. Working with a single computer or input device necessarily limits the extent to which each person can contribute to a computer-mediated task, and it is not difficult to see why this could lead to inequity in contributions. However, studies in this area seem to lack a clear theoretical distinction between *equitable participation*, which seems to be characterised as 'everyone taking part', and the sort of equity in participation that might be characterised by 'everyone pulling his or her weight', i.e. in compliance with fairness norms. Some work in this area is explicitly directed towards this latter conceptualisation (e.g. Plonka *et al*., 2012), potentially making this topic relevant to fairness. Yet many researchers do not provide a clear indication of what equity is intended to mean in their studies (e.g. Harris *et al*., 2009; Wallace *et al*., 2013) making it hard to specify the potential relevance of this literature to the present thesis.

In other areas, fairness arises less in terms of work completion but more with respect to moral or other social issues raised by technology. For example, O'Neill & Martin (2013) argue that crowdsourcing platforms, which facilitate mass-scale task distribution (or so-called 'microwork'), must attend to the rights and potential benefits offered to workers. O'Neill & Martin conducted an ethnographic investigation of two companies responsible for processing healthcare data, finding that seemingly menial tasks carry hidden subtleties and opportunities for, albeit modest, career progression—such aspects could be lost if the work were transported to a crowdsourcing platform. The implication here is that it may be unfair to erode the basic opportunities for personal development associated with even the most innocuous of work tasks. The authors go on to argue that employers of crowdworkers should place greater importance on developing two-way relationships with their anonymous employees. Such relationships should be founded on trust and fairness principles, e.g. acceptable payment for the work done and the ability to provide feedback about an employer's processes.

A final area in which fairness has been identified as relevant to HCI is in *Value-Sensitive Design*, which refers to a "theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process (Friedman *et al*., 2006, p. 349). The definition of a value is broad: "what a person or group of people considers important in life" (Borning & Muller, 2012, p. 1125). Several such values, as identified by prior research, include freedom from bias, trust, and accountability (Friedman & Kahn, 2003). There appears to be no reason why fairness might not be considered as one value that is also potentially relevant to design (Hochheiser & Lazar, 2007), although, as noted above, no work has yet addressed the issue explicitly. Prior work does suggest that fairness is relevant: in an evaluation of a collaborative web search tool, Morris & Horvitz (2007) found that participants did not use features meant for automatic division of labour because only one member of the team had control of the functionality. This distribution of control implied an unfair status difference that was contrary to the collaborative ethos of the group (Morris & Horvitz, 2007). Such a simple design decision clearly had profound effects on the way the system was used.

2.5.3 Summary and Implications

As evidenced by this rather sparse review, there is a clear lack of work on fairness in CSCW and, given the potential relevance of fairness for collaboration, we see an opportunity to assess the extent to which fairness might be desirable in collaborative work and technology design. The aforementioned study by Galegher & Kraut (1990) provides a hint in this regard—groups found it difficult to monitor collaborative contributions in computer-mediated work, and such a problem might remain true in the present day. Indeed, similar arguments about the lack of feedback about contributions in distributed, technologically-mediated settings have been made elsewhere (e.g. Hertel *et al.*, 2002, 2003).

For the present thesis, we aim to study fairness in the division and completion of collaborative work. We suggest fairness is relevant in two ways. First, in guiding the initial allocation of tasks around a group. This may not be true in *all* settings, but could certainly apply to those where prescription of tasks is unguided and the distribution of control is relatively equal. Second, we expect that fairness is relevant to the completion of work. The very existence of the free-rider problem speaks to this contention, but ensuring fairness in completion may be especially difficult when a group's work is mediated by technology—such situations are often characterised by limited awareness and accountability of contributions. In the following chapter, we aim to develop these ideas by investigating fairness in the context of collaborating student workgroups. We explore the extent to which tasks are distributed in accordance with fairness norms, and whether non-participation arises as a concern. This helps to provide further justification for the work reported in later chapters.

2.6 Chapter Summary and Conclusion

In this chapter we selectively reviewed literature relevant to the work presented later in this thesis. We first explored the notion of computer-supported cooperative work (CSCW), and then focused on understanding the nature of collaborative work itself. We reviewed several existing attempts at mapping out concepts relevant to collaboration, and acquired a basic definition of collaboration appropriate for this thesis. We then discussed the concepts of coordination, common ground, awareness, and division of labour, highlighting why each is relevant to the design of CSCW systems. This led to our consideration of fairness and its role in human society. We then reviewed prior studies where fairness has been mentioned in CSCW, and finally articulated our own interests for the present thesis. The following chapter describes thesis study 1, where we explored fairness and its relation to satisfaction with group outcomes in small groups of 5–6 undergraduate students.

Chapter 3

A Study of Division of Labour and Fairness in Collaborating Teams

3.1 Chapter Overview

In Chapter 2 we raised the issue of fairness as a potential topic of concern for collaboration. This chapter presents an exploratory study of division of labour and fairness in small collaborating workgroups. Specifically, we use a survey instrument to quantify perceptions about fairness in division and allocation of work, and explore how these perceptions are related to three dependent variables: the overall satisfaction with the products of each group's work; the work process; and the performance of team members. The workgroups in question were comprised of undergraduate students who had recently completed year-long collaborative software development projects. Marks from these projects contribute to each student's final degree classification, meaning that all team members had a vested interest in ensuring that the work is competed to a high standard.

The purpose of the study is to obtain some quantitative, non-experimental evidence that implies the potential importance of fairness during group work. We explore the extent to which fairness and satisfaction are correlated—one might imagine that, if fairness were important, individuals who perceived that their work was completed in line with fairness norms would show greater satisfaction than those who suffered because of free-riding or non-participation from team members. Additionally, we gather initial qualitative data that describes challenges faced by students and the extent to which problems encountered are related to fairness and non-participation from team members. Such responses also highlight the sorts of considerations that are taken into account when managing division of labour in groups. The results of this study validate the concerns of the present thesis, providing evidence of a) a relationship between perceived fairness in the division and completion of work, and satisfaction with group outcomes, and b) the difficulty of monitoring and gauging equality among team member contributions during computer-mediated collaboration. We also consider the extent to which satisfaction might be influenced by individual control, team member abilities, and group construction.

3.2 Thesis Study 1: Survey Study

3.2.1 Study Motivation

As mentioned above, the present study sought to explore the potential links between fairness, division of labour, and satisfaction in the context of real-world teams. Our chosen teams were small groups of students, each comprised of 5–6 individuals, who had recently finished collaborating on year-long software development projects (known as the Integrated Project or 'IP') at the University of Bath. Students in these groups are assigned to teams by the unit lecturer at the outset of the academic year. Groups must then work to design, implement, and evaluate a complete software system, usually taking the form of a website or mobile application. Students must also submit four paper-based deliverables documenting their progression through various stages of the software development lifecycle.

Our selection of these students was partly one of convenience, but can be justified in terms of the highly collaborative nature of their projects and the importance of the work itself—marks from the finished software product and its associated documentation contribute towards the final degree classification of each team member. Students should, therefore, have been interested in creating projects that are of the best possible quality, and these projects always entail division of labour among the team. Additionally, the author's first-hand experience of tutoring groups on the IP unit suggests that students do experience negative emotional consequences when team members fail to contribute good work or refuse to participate altogether. In our experience, students are quick to grasp an opportunity to express dissatisfaction with non-participating team members. These considerations gave us good reason to believe that fairness would be a foremost concern for students and could be studied in relation to different aspects of project completion.

In the present study, we used a survey instrument to explore students' perceptions about division of labour, fairness, and overall satisfaction with the outputs of their group. Our survey was administered as a one-off exercise to students after each group had submitted their final software system and paper-based deliverables, but before each individual had received their final marks for the unit. We opted to use this critical project interval because the experience of collaboration was still fresh in students' minds, but was not yet coloured by their final marks. We believed that this would minimise the potential negative impact of low or dissatisfactory marks on retrospective evaluation, as could be the case if the survey had been sent *after* students had received their final grades.

3.2.2 Survey Design

Our survey was designed to acquire quantitative and qualitative data about a range of issues associated with group work. First, and in line with the concerns of this thesis, we wanted to acquire insight about fairness in students' division of labour. We were interested in any problems students encountered in managing divided work, especially those related to fairness and non-participation from group members. Second, we were interested in perceptions about fairness in the division *and* completion of work. As we have seen, the former pertains to initial allocations of work, which we might expect to be taken in line with equality in the relatively egalitarian settings of student workgroups, but could also be adjusted in line with individual competencies (making fairness the appropriate term in this case). Conversely, fairness in completion pertains to the way in which work was actually carried out—it may be the case that perceptions about this stage will be different if groups incur free-riding behaviour. We aimed to delineate these constructs by using separate measures for each within our questionnaire.

Third, we wanted to explore the impact of fairness, in both division and completion of work, on overall satisfaction with group outcomes. We operationalised three measures of satisfaction: satisfaction with the *product* of the team's work, i.e. the system and its documentation; satisfaction with the *process* used to assign work, i.e. the chosen method of achieving division of labour; and satisfaction with the *performance* of team members, i.e. whether or not team members contributed to an acceptable standard. We reasoned that all three of these constructs might be influenced by perceptions about fairness in the division of labour. Our focus on the concept of satisfaction is guided by its use as a dependent variable elsewhere in the HCI literature (e.g. Galegher & Kraut, 1994; Oulasvirta *et al.*, 2009) and because it is a neutral term that avoids emphasizing particular emotions while, nonetheless, offering a link to happiness and fulfilment (Tatarkiewicz, 1976). Moreover, it is used prevalently in psychological and personality measures, e.g. the much-cited 'satisfaction with life scale' (Diener *et al.*, 1985).

3.2.2.1 Scale Development

We developed a series of Likert-type scales intended to gather quantitative data about students' perceptions of division of labour, fairness, and satisfaction. Table 3.1 lists the items that comprise each of the scales used in our survey. After surveying the literature for existing scales on division of labour and fairness, we found that a prior study by Galegher & Kraut (1994) reports scales used to measure perceived fairness in contributions. However, their scale was intended to be used on a daily basis, and was phrased using the present tense (e.g. *"We are all contributing fairly to this project"*), whereas our survey was a one-off exercise that called for retrospective phrasing. We were not able to acquire an established measure after further literature review, and thus we used the work of Galegher & Kraut (1994) to inform our scales for gauging perceptions about fairness in division of labour. Similarly, our scales of satisfaction were informed by those of Galegher & Kraut (1994) and by the Group Satisfaction scale, an established measure reported by Hackman (1988). Finally, we included a single-item scale to gauge perceptions of equality in contributions to the common project. This was done to allow for comparisons between our two measures of fairness and a straightforward measure of general equality.

Additionally, we incorporated several other established measures of group attributes into our questionnaire. Not only did this serve to mask our primary interest in division of labour and fairness within the questionnaire, it also allowed us to explore whether our concepts of interest were related to other measures reported within prior literature. We identified five existing scales used to study the effectiveness and stability of small groups:

Measures of Fairness and Satisfaction
Fairness in Division of Work, 3 items
My team made an effort to ensure that work was distributed fairly among members
In my opinion, the division of work among my group members was fair.
*When dividing up tasks, some members of my group were given an unfair amount of work.****
Fairness in Completion of Work, 4 items
I believe that, by the end of the project, everyone completed a fair amount of work.
*One or more people in my group did significantly more work than others.****
Everyone contributed an amount of work that I consider was fair.
*Compared to the average amount done by other group members, I did an unfair amount of work.****
Satisfaction with Products from Work
Overall, I am satisfied with the quality of the reports our group produced.
Overall, I am satisfied with the quality of the system our group produced.
Satisfaction with Process of Assigning Work
I am satisfied with the way in which work was allocated among group members.
Overall, I am satisfied with our chosen method of allocating work to the members of our group.
Satisfaction with Performance of Team Members
The quality of work produced by the other members of my group was satisfactory to me.
Overall, I am satisfied with the collective performance of my team members.
Equality in Completion of Work
All members of my group contributed similar amounts of the various subtasks required by the assignment.

Table 3.1: Scale items used to gauge fairness and satisfaction in survey, study 1. Reverse scoring is denoted by ***.

Open Communication, 4 items: measures the extent to which group members feel that they are able to speak their minds. Stokes (1983) argues that cohesive groups are more likely to take risks that lead to useful and novel outcomes. To take those risks, group members must be able to speak freely and make statements that may be counter to current opinion.

Task Motivation, 4 items: gauges perceptions about whether a given team is task-oriented and focused on work (Zaccaro & McCoy, 1988). High task motivation is seen as a positive for a group.

Group Viability, 3 items: assesses how well a group functions as a team and whether team members perceive their experience as positive (Hackman, 1988).

Group Cohesion, 5 items: this set of measures assesses the extent to which individuals show affinity and a liking for their teammates (Stokes, 1983), both of which are taken as a sign of good cohesion.

Social Loafing, 3 items: quantifies perceptions about social loafing behaviour and the non-participation of particular individuals within a group (Druskat & Wolff, 1999).

The items that make up each of these scales are shown in Table 3.2. We believed that several of

Measure
Open Communication, 4 items (Stokes, 1983)
*My group avoids saying anything that might upset someone.******
My group is very straight-forward with me.
*There are certain topics about our work that my group avoids talking about.******
*Most people in my group are careful not to reveal too much of themselves to the group.******
Task Motivation, 4 items (Zaccaro & McCoy, 1988)
Performing well is a top priority for my team.
My group members expect high effort and commitment from me.
Only a high level of performance is acceptable to our group.
Our group is highly task-oriented.
Group Viability, 3 items (Hackman, 1988)
*As a team, this work group shows signs of falling apart.******
Members of my team care a lot about it, and work together to make it one of the best.
Working with members of my team is an energizing and uplifting experience.
Group Cohesion, 5 items (Stokes, 1983)
If I were to participate in another group like this one, I would want it to include people who are very similar to the ones in this group.
*Most of the people in the group are not the kind of people I would enjoy spending time with outside the group sessions.******
*There are not many people I like as individuals in my group.******
Even if we stopped meeting as a group, I would still want to see the people in this group as often as I could.
I wish I had more time for socializing with other group members.
Perceived Social Loafing, 3 items (Druskat & Wolff, 1999)
We have some team members that dont put much effort into their work.
*Every member of our team does his/her share of the work.******
There are some individuals on our team who dont do much work.

Table 3.2: Established scale items used in survey, study 1. Reverse scoring is denoted by ***.

these constructs could be meaningfully related to fairness and overall satisfaction; for example, one might imagine that cohesive and well-functioning groups might work in line with fairness norms, and that fairness and satisfaction might be low if social loafing is high. Use of these scales also allows us to obtain a broader perspective on the overall working health of our groups of interest.

Following the work of Galegher & Kraut (1994), all measures in Tables 3.1 and 3.2 were implemented using 7-point Likert scales, with 1 = strongly disagree and 7 = strongly agree.

3.2.2.2 Free-text Responses

Our questionnaire included spaces to gather qualitative data concerning participants' experiences. Specifically, we asked: *Do you have any other comments or anything else you would like to add regarding your satisfaction with the Integrated Project?*, and *Is there anything else you would like to add about your*

Integrated Project group that was not covered by the survey?. We also gave participants the opportunity to expand upon, and clarify, their ratings on each survey page. All free-text responses were optional so as not to overburden respondents and to avoid increasing dropout rates.

3.2.3 Survey Pilot

Our preliminary questionnaire was piloted using a small sample of postgraduate computer science Master's students who had recently completed a coursework assignment in groups of 3–4 individuals. Respondents were solicited directly via email, and eight people completed the questionnaire. Initial results suggested correlations between our variables of interest, and the study also allowed us to make adjustments to the wording of our own scales after some participants expressed confusion or misunderstanding. Final scales were identical to those shown in Tables 3.1 and 3.2.

3.3 Method

3.3.1 Participants

Respondents were second year students enrolled in the 2011–12 undergraduate Computer Science course at the University of Bath. As mentioned above, all respondents had recently completed a group software development project, which required the group to design, implement, and evaluate a working software system. From a possible 68 students in the class, 38 provided a response (56% of class). Ten respondents were female and 28 were male. Respondents' ages ranged from 19–32 years (*Mean* = 20.6, *Median* = 20).

For the purposes of the Integrated Project, the class was separated into 12 subgroups. While all groups were represented within our data set, we did not receive an equal number of responses from each group. (Lower boundary = two responses, upper = five responses).

3.3.2 Survey Implementation and Procedure

The survey was hosted online using Qualtrics survey software.[4] To reiterate, all measures in Tables 3.1 and 3.2 were implemented using 7-point Likert scales, with 1 = strongly disagree and 7 = strongly agree.

The survey had six pages in total, and was designed to be completed quickly because, based on our personal experience of collecting student feedback, it can be notoriously difficult to get undergraduates to complete evaluation questionnaires. (A full transcript of the survey can be found in Appendix B). Participants were first asked to provide informed consent and were told that all of their responses would be stored securely and confidentially. We stressed that responses would not affect their final mark for the unit. The second page of the survey requested demographic information including age and gender. The third page contained the group construct measures, presented in a different randomised order for each participant to prevent order effects. The fourth page gauged perceptions about fairness in completion of

[4]http://www.qualtrics.com

work and assessed the assignment of work among group members, and the fifth page assessed satisfaction with group products, performance and procedure. Again, all scales were presented in a randomised order. The final page thanked the student for their participation and offered them the opportunity to leave their username to be entered into a prize draw. Usernames were stored separately from responses, meaning that we could not directly identify any individual. Students were made aware of this in the survey.

After the survey was hosted online, two electronic mailouts were made inviting students to complete the survey. These mailouts were sent 3 days after students had submitted their final assignment. Ten prizes of £10, drawn and awarded randomly to those who completed the survey, were offered to encourage responses. The survey was closed after a third mailout failed to elicit any further responses. Prize winners were selected at random by an independent individual.

3.3.3 Analysis

For the present study we opted to perform analysis at the level of individual responses. While it would be desirable to also explore our data using groups as the unit of analysis, such an approach would be statistically unsound as we do not have equal numbers of responses from each group. Our analyses focus on the interpretation of correlations using non-parametric statistical tests. While a regression analysis might be useful to estimate the relationship between fairness and our various measures of satisfaction, such an approach would be weakened by our low number of participants and the fact that our criterion variables, i.e. measures of satisfaction, are not normally distributed. These issues mean that we cannot meet the basic criteria for regression (Dancey & Reidy, 2007; Tabachnik & Fidell, 2007).

In line with similar analyses in prior work (e.g. Druskat & Wolff, 1999; Foo *et al*., 2006), we obtained composite scores for each participant in the study; such scores are obtained by computing the mean of an individual's responses on a given subset of items (e.g., the average of all four items on the Open Communication scale). Mean scores are easier to interpret because they are constrained within the original metric and can be understood within the confines of the original scale (Griffin, 2009). Composite scores were used for all of our analyses, unless otherwise stated.

The internal reliability of each of our scales was assessed using Cronbach's alpha. For scales with only two items, the Spearman-Brown stepped-up reliability coefficient, identified as the most appropriate consistency measure for 2-item scales (Hulin, 2001), was used to gauge internal reliability. All scales were found to have high consistency, bar Open Communication with an alpha of .58. This value is similar to that observed by Druskat & Wolff (1999) (.56 in their study, where $n = 400$) and perhaps suggests that the potential relevance of this scale should be interpreted with caution. Since it shows only weak to moderate correlation with the rest of our measures of interest, we will not present any in-depth considerations of the measure, but will nevertheless retain the correlations in our results set because they do not harm our analysis.

	Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11
1.	Openness	4.56	1.02	(.58)										
2.	Task Motivation	5.14	1.27	.36*	(.89)									
3.	Viability	4.66	1.47	.43**	.74**	(.84)								
4.	Cohesion	4.30	1.19	.51**	.62**	.79**	(.82)							
5.	Social Loafing	4.64	1.80	–.33*	–.46**	–.64**	–.59**	(.91)						
6.	Division Fairness	5.13	1.08	.44**	.45**	.49**	.49**	–.47**	(.74)					
7.	Completion Fairness	3.38	1.47	.38*	.45**	.69**	.62**	–.82**	.64**	(.89)				
8.	Product Satisfaction	5.29	1.66	.38*	.80**	.66**	.58**	–.35*	.40*	.35*	(.91†)			
9.	Process Satisfaction	5.04	1.35	.58**	.69**	.76**	.67**	–.51**	.65**	.61**	.67**	(.94†)		
10.	Performance Satis.	4.91	1.66	.40*	.82*	.78**	.64**	–.64**	.56**	.57**	.81**	.79**	(.91†)	
11.	Contribution Equity	3.47	1.95	.25	.57**	.69**	.52**	–.52**	.47**	.66**	.44**	.64**	.59**	(n.a)

Table 3.3: Correlations among variables for entire sample (n = 38) and observed reliability coefficients for individual scales.
Note: Cronbach's alpha shown along the diagonal in parentheses where appropriate. Two-item scales tested using Pearson-Brown coefficient, denoted †.
* = $p < 0.05$ — ** = $p < 0.01$

3.4 Results

3.4.1 Correlational Analyses

Table 3.3 shows the strength and direction of correlations between measures used in our survey. Since data were non-parametric, and scores were not normally distributed on several scales, all correlations were computed using Spearman's rho. In general, correlations are of moderate strength, and all of the correlations are significant, with p <0.05 throughout. For our primary topics of interest—fairness and satisfaction—we can see correlations between both aspects of fairness and all three measures of satisfaction. Additionally, Table 3.3 shows that all three measures of satisfaction are positively correlated with one another. This is perhaps not surprising; one might intuitively expect that those who were quite satisfied with the process and products would also be satisfied with performance of team members.

Below we focus on particular correlations of interest and later sketch some potential interpretations in our discussion section. Our discussion section also reflects more broadly on the possible meaning of our results, as well as limitations.

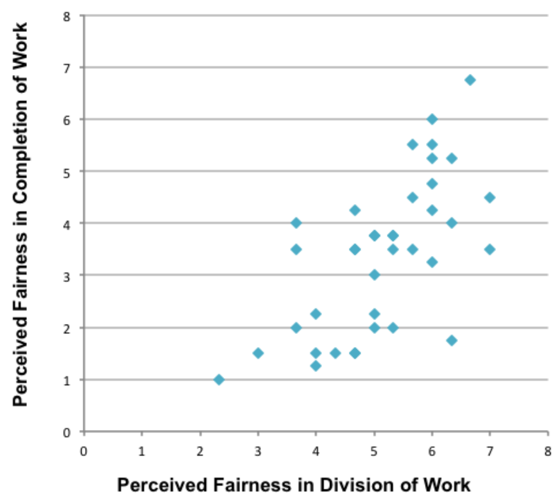

Figure 3.1: Scatterplot illustrating correlation between perceived fairness in division of work and perceived fairness in completion of work.

Fairness in Division and Completion of Labour

Table 3.3 shows a positive correlation between fairness in division of work and fairness in completion of work, r(37) = .64, $p < 0.01$. This may be indicative of a dual fairness trend—if groups operate in line with fairness at the outset, perhaps the established norm follows through into the completion work as well. However, comparison of the means in Table 3.3 suggests that, although most individuals perceived their team as having made an effort to ensure that work had been divided fairly (M = 5.13), the average for fairness in the *completion* of work (Row 7) is slightly lower (M = 3.38). This difference is significant, paired t[37] = 9.47, $p < 0.001$. Eyeballing the data in Figure 3.1 indicates that some individuals in the lower right area of the chart perceived high initial fairness but rated completion much lower. This suggests that some initially fair allocations may not have panned out as intended. Of course, joint projects often necessitate reshuffling of workloads, but it is also possible that the decrease is due to free-riding behaviour within our groups of interest.

Fairness and Process Satisfaction

Table 3.3 shows that satisfaction with the work *process* has a moderately strong, positive correlation with perceived fairness in the division of work, r(37) = .65, $p < 0.01$. The relevant data are graphed in Figure 3.2 and illustrate the expected relationship between fairness and satisfaction—overall, those who believed work had been divided fairly also express high satisfaction with this approach.

A similar correlation exists between process satisfaction and perceived fairness in the completion of work, r(37) = .61, $p < 0.01$. It is worth noting that our measure of process satisfaction pertains to the *initial* assignment of work, whereas fairness in completion gauges perceptions about what occurred while work was actually being completed. This perhaps makes this correlation somewhat less direct.

A more interesting finding is the correlation between fairness in completion and perceived equality in contributions, r(37) = .66, $p < 0.01$. The general trend is illustrated in Figure 3.3, although it is worth noting the presence of some individuals who perceive low fairness but report reasonably high equality and, conversely, those who perceive reasonably high fairness but low equality. It is possible that these results stem from competency based workload assignments—again, we will return to these considerations in our discussion.

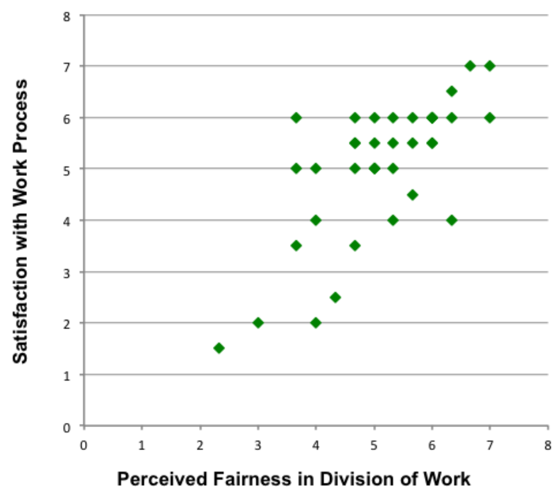

Figure 3.2: Scatterplot illustrating correlation between perceived fairness in division of work and process satisfaction.

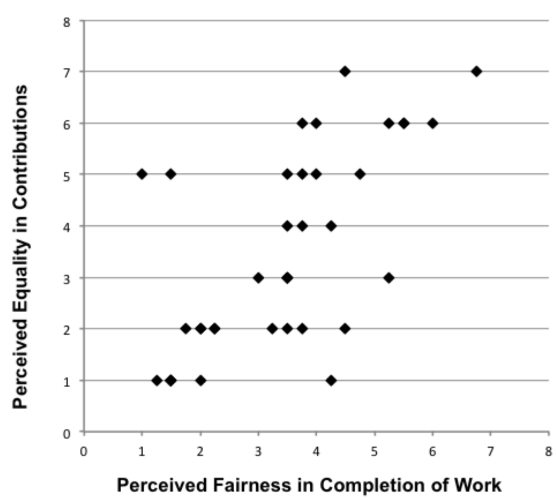

Figure 3.3: Scatterplot illustrating correlation between perceived fairness in completion of work and perceived equality in contributions.

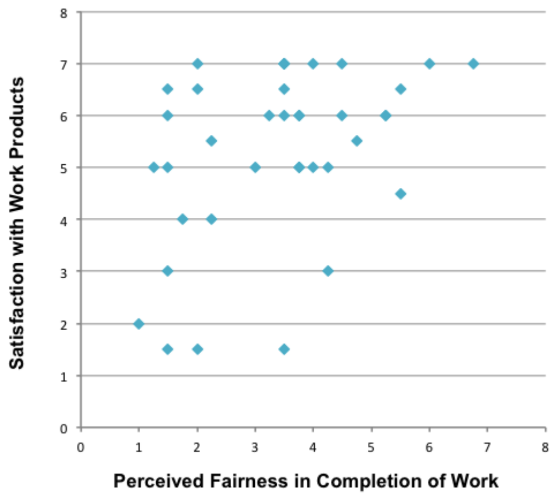

Figure 3.4: Scatterplot illustrating correlation between perceived fairness in completion of work and satisfaction with work products.

Fairness and Product Satisfaction

Turning to Product Satisfaction, the correlations with Fairness in Division, $r(37) = .40$, $p < 0.05$, and Fairness in Completion, $r(37) = .35$, $p < 0.05$, are noticeably weaker than that of Process Fairness. Of these two correlations, fairness in completion is likely to be more relevant to product outcomes, and the relevant data from this correlation are shown in Figure 3.4. It can be seen that some individuals, towards the upper left of the chart, believe that work was not completed fairly yet still appear to be very satisfied with the group outputs. Of course, we can only speculate as to why this may occur, but some plausible explanations can be found within our qualitative data—for now we defer interpretation of these results until our discussion section.

Fairness and Performance Satisfaction

Satisfaction with performance of team members shows a moderate positive correlation with fairness in division of work, $r(37) = .56$, $p < 0.01$, and completion of work, $r(37) = .57$, $p < 0.01$. This aligns with the other correlations and suggests a relationship between perceptions about fairness and perceptions about team member performance.

Group Construct Measures

Table 3.3 also displays some notable correlations between our variables of interest and the pre-established measures of group phenomena. In particular, the construct of Task Motivation is strongly and positively correlated with all three measures of satisfaction. This is likewise true of Group Viability and Group

Cohesion, and although the correlations are weaker in these cases, they do remain above moderate strength. Again, we cannot infer causation but it appears that groups who work well, are cohesive, and remain viable seem to achieve better outcomes, as characterised by higher satisfaction across all three measures.

The constructs of Task Motivation, Group Viability, and Group Cohesion are moderately correlated with Fairness in Division of Work. The correlations are slightly stronger for Fairness in Completion of Work, at least for Viability and Cohesion. It is possible that the scales for these latter constructs have a more meaningful relationship to the enactment of work, e.g. *"Members of my team care a lot about it, and work together to make it one of the best"*, in turn explaining the stronger association.

Lastly, Social Loafing negatively correlates with every other measure, particularly Performance Satisfaction. It is not difficult to see how these two constructs might be related. While we cannot make sound claims about causation, the strong negative correlation between fairness in completion of work and perceptions about social loafing ($r(37) = -.82$, $p < 0.01$) suggests a relationship between the two and clearly speaks to the presence of free-riders within our groups of interest. It is also interesting that the correlation between Fairness in Division and Social Loafing is noticeably weaker, $r(37) = -.47$, $p < 0.01$). This perhaps speaks to the possibility that allocations were fair at the outset of work, and later became unfair during actual enactment. (As hinted in our earlier analysis of fairness in allocation and completion.)

3.4.2 Qualitative Responses

Seventeen respondents provided qualitative data. The majority of responses were to the question *Is there anything else you would like to add...?* and so we chose to consolidate responses into an aggregated dataset. Responses were analysed using thematic analysis (Braun & Clarke, 2006) although the analysis was fairly lightweight due to the limited amount of material within the dataset.[5] Our analysis aimed to inductively elicit topics relevant to division of labour, fairness, and group work in general. Transcripts were read twice and initial themes were developed. We then passed these themes along with the transcripts to an independent coder for verification. This resulted in four relevant themes: *assignment of work by competencies*; *social loafing and non-participation*; *concerns about quality of work*; and *awareness difficulties*. A fifth theme, 'Other', was used to classify anything that did not fit our other themes, e.g. opinions about the content of the project unit or facetious remarks. Statements were then classified by the same coder and a second independent individual, with two disagreements about categorisation resolved through discussion. Due to the limited quantity of data (responses were relatively short in length, ranging from one sentence to two paragraphs) we do not make claims about prevalence and prefer to consider responses as illustrative of the types of problems students encounter during group work. Themes are described below using quotes appended with participants' gender and age in the form [M/F, Age].

First, *assignment of work by competencies* pertained to statements mentioning how subtasks were divided in accordance with skills and abilities. Four students mentioned that their group had adopted such an approach. Two mentioned that this served to legitimise inequitable divisions of labour as fair, as

[5] Approximately 1–2 paragraphs per response—later studies in Chapters 6 and 7 involve analysis of larger interview datasets.

illustrated by the following response:

> *"The amount of work done by members was very uneven (eg some did a lot more, some did very little) but it was not always completely unfair in every case. for example the programming pros did a lot of programming, and the others helped out and learned some but did not do as much as the pros on programming."* [F, 20].

Second, *social loafing and non-participation* pertained to non-participation from team members. Five respondents mentioned this as a problem within their groups. In some cases particular team members had dropped out from the course, but in others loafing was related to simple lack of effort:

> *"The main problems related to getting the work done quickly enough and nobody putting a great deal (although some amount was) of effort into the work."* [M, 23].

Third, *concerns about quality of work* pertained to statements where individuals mentioned poor standards of work from other team members, which in many cases necessitated rewriting of the work. Five individuals mentioned quality control as an issue. For example:

> *"I did most of the work in nearly all the deliverables, but I felt too bad to lower anyone else's percentage on the contribution forms. Some members did awful work, but it was more convenient for me to simply redo it than to try and get them to improve."* [M, 20].

Finally, *awareness difficulties* pertained to the difficulty of gauging individual efforts during highly subdivided projects. This may have affected perceptions about fairness to the extent that lack of visibility meant people were more lenient on suspected free riders. Three individuals mentioned this:

> *"Contributions were fairly specialised by subtask, so the amount of work I saw people doing was obviously a lot higher amongst the people who were working closely with me. It is difficult to see what others are doing most of the time, but I still believe everyone contributed a fair amount to the project as a whole."* [M, 22].

We regard these themes as offering insights into the issues that may have shifted perceptions about fairness within groups—we use other quotes from our themes to complement our interpretations below.

3.5 Discussion

The present study sought to explore the relationships between division of labour, fairness, and satisfaction via a survey of small collaborating workgroups. We found that perceived fairness correlated with several measures of satisfaction. Here we attempt to sketch interpretations and elaborate on other factors that could influence satisfaction and decisions about fairness in division of labour.

Our foremost finding was a positive correlation between fairness in division of work and satisfaction with the process of work assignment. Such a finding is a very direct demonstration of the relationship between fairness and satisfaction, and is in line with our initial contentions about the potential importance of fairness in division of labour. More tentatively, the finding perhaps indicates that fairness is the guiding workload allocation of choice within groups. However, we must be careful about our claims given the nature of our data, which is based solely on retrospective self-reports—we cannot definitively state that increased fairness also leads to increased satisfaction. This is because, although fairness in division of labour could conceivably lead to greater satisfaction, the reverse is also true in that participants' memories about how work was divided could be influenced by their general feelings about the project as a whole. We would stress that more work is required before any definitive conclusions can be drawn about the relationships we have identified.

A further finding was that perceived equity in contributions was positively correlated with perceived fairness in completion of work. This speaks directly to the potential desire for roughly equal contributions, with those reporting low equality also reporting low satisfaction. However, it is worth noting that two individuals rated perceived fairness as moderate, but equality as low (fairness$>$5, equality $<$2.5, see Fig. 3.3 above). These ratings could be associated with the assignment of work by competencies—although there may be some basic expectation for equality, it is not uncommon for the sorts of student workgroups surveyed here to subdivide and assign tasks based on expertise, e.g. according to skill at computer programming. A qualitative response provided by one of the individuals in question speaks to this approach:

> *"We split the written aspects of the coursework up fairly evenly, but we split the programming component amongst three of the six of us. We decided as a group that this was the best way to do it. Generally to compensate for this additional workload, the more complex and time-consuming parts of the non-programming parts of the coursework were given to the three who weren't doing the programming; e.g., the design aspect of the presentation."* [M, 23].

Such an approach might be regarded as fairer but could lead to inequality given that programming comprised a relatively large part of the students' assignment. Conversely, some individuals perceived fairness as low but equality in contributions as high. The individuals in question did not leave qualitative responses to explain their ratings, but one possibility is that their groups may have chosen to divide equally without taking competencies into account. Such an approach could cause certain individuals to struggle with the work, potentially leading to the perception of unfairness. These are, of course, speculative intuitions, but are nonetheless realistic explanations as to why disjoints between equality and fairness might have occurred.

We also found that fairness in division of work was correlated with fairness in completion, potentially implying that initial fairness feeds through into the completion of work. However, this was not true for all respondents—data in Figure 3.1 suggest that several participants believed that work was initially allocated on the basis of fairness but ended up unfair by the time of completion. Of course, joint projects often necessitate the reassignment of workloads, but participants' qualitative statements suggest that some groups suffered from free-riding behaviour, in turn offering a potential explanation for the perceived drop in fairness. For example, one individual (fairness in division = 5, fairness in completion = 2) offered the following explanation for his ratings:

> *"One of our group members left our group a few weeks into the project - therefore our initial work distribution needing changing. Another issue when distributing work was that there was at least one team member that we were not entirely satisfied with. And as the project progressed, we became less trusting of this person's work and tried to assign tasks that were not overly critical to the progress of the project. This wasn't the best "group work" approach, but if this person had been given more work, it would have fallen to myself and another team member to go over the work to get it up to the level that we expected and that we would be happy to hand in for assessment."* [M, 20].

The presence of a correlation between social loafing and fairness in completion of work further emphasizes this issue, with those reporting high social loafing also reporting very low fairness in completion.

While one might be inclined to suspect that low fairness in completion of work has a universally negative impact on outcomes, this may not have been the case for all of the individuals we surveyed. Our analysis of product satisfaction (see Fig. 3.4) indicated that five individuals perceived low fairness but were, overall, highly satisfied with their group's collective outputs (n = 5, fairness $<$3, satisfaction $>$5). Although we would stress a need to collect further data on this issue, several potential explanations are available. First, high product satisfaction in the face of unfairness may again be related to the variance in individual competencies within our groups of choice. The team structures in our student sample were externally-imposed, meaning that individuals may suffer the effects of being placed with others who have little interest in the project. In such cases, non-participation from specific team members might mean that the remainder of the group has greater control over the project's content. In turn, individuals who have contributed more might be unhappy about free-riding but would be more satisfied with a project that had a greater percentage of their own work within it. A statement from respondent 3 (perceived fairness = 1.5, product satisfaction = 6) speaks to this: *"One member gave awful quality work, and I redid all of it"* [M, 20]. An alternative explanation is that individuals might be more satisfied with products that arise out of adversity. For example, the greater effort investment required to overcome the effects of free-riders might skew satisfaction with group outputs, leaving individuals feeling the 'warm glow' of success (cf. Isen, 1970). Our limited dataset prohibits us from exploring these issues in detail, but each could conceivably be explored in further work.

Our final correlations of interest were those between established measures such as Task Motivation, Group Viability, and Group Cohesion. In general, these measures were strongly and positively correlated with measures of satisfaction. These findings might be regarded as further evidence suggesting that high scores on these measures are generally associated with better group outcomes (Stokes, 1983; Hackman, 1988; Zaccaro & McCoy, 1988).

Although we have observed correlations between fairness and other variables of interest, some of the correlations could be problematic if our measures were used as predictors in a regression model. We initially anticipated that we would use regression in our analysis but could not because our data violated several of the assumptions that underly regression models. One such assumption is the absence of *multicollinearity*, which occurs when there is high intercorrelation among predictor variables. Such intercorrelations violate several of the assumptions that underly the mathematical estimation of regression models. The best regression situation thus occurs when predictor variables correlate with criterion variables but not with each other (Dancey & Reidy, 2007). Almost all of the explanatory variables in our study were correlated, e.g. fairness in division and fairness in completion (r = .64) and social loafing and completion fairness (r = -.82), implying multicollinearity.

We do not believe that multicollinearity is a problem for the present study because we did not perform regression. It would, however, be prohibitive to more advanced analyses in future work. One common approach to resolve multicollinearity is to set an arbitrary threshold for correlations among predictor variables and then eliminate unsuitable predictors from the model. For example, a study by Reed *et al*. (1985), which examined the relationship between diet and blood pressure in men, used $r >$ .7 as the criterion for elimination. Predictor variables that were correlated (e.g. milk and calcium) were evaluated using separate regression analyses. Other authors (e.g. Mansfield & Helms, 1982; Frost, 2013) suggest examination of variance inflation factors and standardisation of predictors as ways of resolving multicollinearity. Neither of these is necessary if one is not performing regression but should be kept in mind if the reader intends to use our study measures in future work. Since almost all of our variables were correlated, future research should perhaps explore fairness and satisfaction in relation to other variables, e.g. the number of persons within the group or the type of work that the group chooses to complete.

Likewise, the correlations between our fairness measures and all three measures of satisfaction are extremely similar; for example, the strength and direction of the correlation between satisfaction with process fairness in division (r = .56) is almost identical to that of fairness in completion (r = .57). This could make it difficult to tease the two apart in a regression model. Future work could benefit from a cleaner distinction between initial fairness and the way in which agreed allocations played out over time. This could be achieved by collecting several readings of perceived fairness during completion of the protracted group assignment. The groups we studied began their group project in October 2011 and completed the assignment in April 2012. In between these dates was a series of deadlines for various subcomponents of each group's project. A future replication could see the researcher take one measure of fairness at the outset and another at each project interval.

3.5.1 Qualitative Findings

We also gathered participants' qualitative responses, which, although somewhat limited in number, were indicative of the sorts of difficulties encountered during divided work. For example, respondents spoke of the need to rework assignments, problems associated with poorly performing or non-participating team members, and of the difficulties with monitoring and gauging contributions during collaborative work. This latter aspect is especially interesting as it hints at a potential opportunity to improve groups' awareness of each individual's contributions to the shared project. This in turn could support judgements about fairness and, presumably, lead to greater satisfaction if individual members can be held accountable for their lack of work.

Overall, our results do suggest that fairness can be a salient concern for collaborating teams, especially when particular team members fail to contribute in line with fairness norms. We also offered tentative insights regarding some of the ways in which fairness preferences might conflict with different aspects of group work. Nevertheless, the fairly limited scope of the present study means that it cannot be regarded as anything more than an exploration of fairness in division of labour—indeed, this is how the present study was initially conceived. The following chapter probes our issues of concern in more depth by introducing an empirical model for exploring division of labour. Chapter 4 introduces the conceptual foundations for our model, and Chapter 5 reports a series of studies where our model is explored in an empirical setting.

3.5.2 Study Limitations and Future Work

Although the collection of quantitative survey data helped to suggest relationships between variables, it obscured the fine details about what occurred within each of our surveyed groups. Placing greater emphasis on qualitative responses would have helped to shed light on some of the issues that arose through our analyses. Also, if we had managed to obtain a complete results set, i.e. a response from everyone in the class, group level analysis might have allowed proper comparisons between groups where free-riding was, and was not, present. The fact that we were not able to get responses from the entire class could additionally mean that our data is reflective of particular personalities, i.e. prosocial types. It is also possible that the people who did not respond even after our three mailouts were the free-riders mentioned by some of our participants.

The Integrated Project unit completed by our respondents is a year-long exercise, and our results are based on students' perceptions captured at a single moment in time. It would be interesting to survey students at intervals during the project to examine how their perceptions about fairness change over time or on a daily basis over a short period of intense work (prior examples of these approaches can be seen elsewhere, e.g. Galegher & Kraut, 1990; Druskat & Wolff, 1999). Collecting data at set intervals may prove pragmatically difficult, however, given that not all groups work at the same speed—the end of the Integrated Project was the only time at which we could be sure that each group's progress was similar.

This work could be extended by comparing ratings of satisfaction and fairness against the marks students were awarded based on their projects. Although our survey results can be interpreted as signalling a preference for fairness in both process and products, we did not ascertain whether those groups with higher fairness actually outperformed those where unfairness was rife. In other words, it would be interesting to explore whether groups with more equal completions of work actually achieve better performance than other groups. An additional extension would be to collect information about the contributions of each team member. While it would be possible for us to gain access to such data by checking contribution forms filled out by the students, we chose not to perform such analyses as we did not have informed consent from our participants.

Lastly, it is important to note that our survey is reliant on students' conceptualisations of fairness—

these are likely to be stable at the individual level but perceptions about what is fair in a team could differ from person to person. For example, those responsible for the programming parts of the project might perceive the increased coding load as very unfair, even if the rest of the team perceives a competency-based approach is much fairer. Thus it is important to bear in mind that perceptions can vary within a group about who is contributing fairly. Further inter- and intra-group analyses would help to explore these issues.

3.6 Chapter Summary and Conclusion

In this chapter we reported a survey study of small student workgroups. Our survey was designed to gauge perceptions about fairness in the initial division of work, as well as in the completion of work during the project. We investigated the extent to which fairness is correlated with satisfaction with a group's work products, the process used to assign work, and the performance of team members. We found that:

- Individuals generally perceived that their group's workload divisions were made in line with fairness, but it appears that fairness in the completion of work was less prevalent. This was likely due to the presence of free-riders within some groups.

- Fairness in the *division of work* was positively correlated with *satisfaction with the process* used to divide work, suggesting a relationship.

- Fairness in the *completion of work* was positively correlated with *equality in contributions*, further implying the relationship between the perception of fairness and a general basis of equality. However, the issue of *competence* muddies this relationship—some individuals rated equality as very low, yet perceived this to be very fair. The reverse was also true, with several individuals reporting high equality but low fairness. It is possible that this relates to individual skills, in the sense that it may be regarded as fair for those skilled at a particular task to complete more of it, hence causing noticeable inequality.

- Fairness in the *completion of work* showed a positive correlation with *satisfaction with products*, though this correlation was somewhat weaker than others in our study. Inspection of the data indicated that some individuals perceived work as highly unfair but were nevertheless highly satisfied. We suggested that this may be related to higher individual control, or possibly the result of a retrospective 'warm glow' associated with overcoming the adversity of unfairness.

- Qualitative responses indicated that some groups did experience free-riding and non-participation, and many spoke to the difficulties of managing divided labour in student teams. We also found that some individuals considered it difficult to monitor the contributions of team members, in terms of who had contributed what to the project. This potentially suggests an opportunity for computer support for fairness.

In the following chapter, we introduce a novel approach for exploring division of labour in an experimental setting. Our approach is based on the classic 'ultimatum game', a model of negotiation that has been used to study division of money in hundreds of prior studies. We adapt the model to explore fairness in the division of workloads.

CHAPTER 4

MODELLING DIVISION OF LABOUR AS AN ECONOMIC GAME: BACKGROUND

4.1 Chapter Overview

In Chapter 3 we found that fairness in the division and completion of work was a salient concern for members of collaborating teams. This chapter offers an alternate perspective on fairness by exploring a model of division of labour based on an economic game. Our game is based on a particular method from behavioural economics. As a field, behavioural economics uses experimental methods to examine the effects of social, cognitive, and emotional factors on the economic decisions of individuals and institutions (Wilkinson, 2008). Among the field's chief contributions is a deep literature that explores fairness in the context of economic decision making (see, e.g., Camerer, 2003). Much of this literature makes use of simple economic games, examples of which include the much-studied prisoner's dilemma (Poundstone, 1992) and the classic public goods problem (Komorita & Parks, 1995). Such games typically require players to make private decisions where individual gain is tempered by negative consequences for the collective good. The fact that games distil intriguing cooperation problems into crisp and simple form is of broader scientific appeal because it allows for the exploration of fairness preferences in a controlled empirical setting (Wilkinson, 2008).

Inspired by the economic literature, this chapter introduces a novel approach to studying fairness in collaboration. Our approach is based on the idea of modelling division of labour in the style of an ultimatum game (Güth *et al.*, 1982). We begin by characterising division of labour as a process of negotiation where collaborators bargain over their individual workloads. We then introduce the classic ultimatum game as a stylized representation of negotiation, and describe how we have adapted its basic structure so as to explore fairness in distributive allocations of workload. The remainder of the chapter provides the groundwork for such an effort by offering a review and critical interpretation of the theoretical and experimental research on ultimatum games. This allows us to consider the various factors that influence behaviour in ultimatum games, such that these factors can be controlled in our studies. The chapter ends by reviewing prior use of economic games in HCI research. We consider how our game can contribute to this ongoing scheme of research.

The contribution of this chapter is the distillation of our model, representing the first attempt at modelling division of labour using an economic game. Our review of relevant experiments guides the methodology employed in Chapter 5, where we report four studies that explore our model using simple collaborative information seeking tasks. The accumulation of literature using economic games in HCI is a further and final contribution.

4.2 Towards a Model of Division of Labour

In this section we consider the nature of division of labour and the extent to which it can be considered as a negotiatory process. This allows us to draw a clear parallel between classical conceptions of division of labour and our adaptation of the ultimatum game.

4.2.1 Division of Labour as Negotiation

As we saw in Chapter 2 of this thesis, division of labour is an integral part of many collaborative work situations and has long been a topic of concern for researchers in CSCW. Numerous authors (e.g., Schmidt & Bannon, 1992; Schmidt, 1994; Eason, 1996; Mark, Haake, & Streitz, 1996) have recognised that collaborative work arrangements often necessitate the assignment of tasks among coworkers. An initial challenge in this regard is to achieve consensus about who is doing what, as well as where, when, and how work will be done (Schmidt & Bannon, 1992).

Multiple scholars (Freidson, 1976; Strauss, 1985; Schmidt, 1994) characterise the assignment of workloads as a process of negotiation where allocations are proposed and then accepted or rejected. This occurs as a back-and-forth process where collaborators repeatedly suggest and evaluate allocations until an agreement is reached (Freidson, 1976). Specific tasks may also be requested, delegated or proffered (Freidson, 1976; Strauss, 1985) according to prescribed job roles (Bardram, 1997), organisational policies and standard operating procedures (Grinter, 1996), or pre-established power structures (Eason, 1996). Yet in an egalitarian setting where no such structures exist, the division of work is likely negotiated (Freidson, 1976; Strauss, 1985; Rogers, 1993; Schmidt, 1994), and it is this basic setting we propose to model and explore here.

As described in Chapter 2 of this thesis, division of labour can be viewed as involving both planning and enacting (cf. Schmidt, 2011), and, in an ideal simplification, these may be considered as two distinct stages (Galegher & Kraut, 1990). In the initial planning stage, collaborators may agree *who* will do *how much* of *what*, i.e., an explicitly planned quantitative division of labour. To prevent redundancy, they may also elect to establish a coordination mechanism, which may also occur through negotiation—deciding not only who will do how much of what, but also *how* it will be done. In the second stage, where work is executed, planned allocations may be followed through, adjusted ad-hoc, e.g. as a result of situated action

(Suchman, 1987; Bardram, 1997) or explicitly renegotiated.

According to these considerations, we focus on two aspects of division of labour. First, the quantitative division of subtasks among participants. This issue arises at the initial planning stage and raises the question of how work should be divided in accordance with skills and abilities, social norms, or local rules established and held by the group. Our second concern is the enactment of agreed workloads, alongside the management of redundancy in joint work results.

In line with our prior discussion of workload assignment, we propose to model the planning and negotiation phase by specifying a simplified version of division of labour in the style of an ultimatum game (UG). The UG is a particular economic game that has been used extensively in studies of negotiation and pecuniary bargaining (Thompson, 1996). The following subsection introduces the game and its theoretical foundations.

4.2.2 The Ultimatum Game: An Empirical Model of Negotiation

The ultimatum game (UG), first studied by Güth, Schmittberger, & Schwarze (1982), is a two-player bargaining game that boils negotiation down to its most basic and final event (Murnighan & Saxon, 1998). In a standard version of the game, players are tasked with dividing a known amount of some commodity, typically a sum of money in the region of US$10.[6] One player is randomly assigned to the role of the proposer and is given an opportunity to suggest a division of the resource. The second player, the responder, then has the option of accepting or rejecting the proposal. If the offer is accepted, the resource is divided according to the proposer's offer. If the offer is rejected, however, neither player receives anything and the money is lost. In a standard, one-shot version of the UG, either of these outcomes ends the game.

Table 4.1 shows the complete range of payoffs for players in a standard UG with a $10 stake.[7] We can see that there are a range of possible options, each of which carries its own degree of risk and reward. The question is, how should players behave if they want to navigate this tricky decision scenario? In the majority of published work on ultimatum games, this question is answered using game theory, a branch of mathematics concerned with modelling the actions of agents in abstracted economic situations, i.e., games (Camerer, 2003). Game theory utilizes the standard economic model of behaviour—which assumes that players in a game are rational, are motivated by utility-maximisation, and are governed by selfish concerns (Wilkinson, 2008)—to determine a formal rule for how a particular game should be played. This rule, commonly known as the 'unique subgame perfect equilibrium' (Selten, 1975), is a prediction that identifies the strategy or set of strategies each player should use if he or she is making the best possible decision while accounting for the decisions of others. The equilibrium is typically derived by backward induction: by working through all possible permutations of the game, one can determine which action the

[6] The dollar amount is given because the majority of published studies have been conducted by American authors. Dollars will be used here to allow for consistent discussion and comparison of the experimental literature.

[7] According to the structure provided in Table 4.1, it is technically possible for a proposer to offer nothing to a responder. However, a responder faced with such an offer has no economic incentive to accept, making the model nothing more than a test of spiteful punishment. For this reason, the minimum offer in a UG is typically set to $1. Table 4.1 simply provides the full range of offers for illustrative purposes.

		P1 Chooses ALLOCATION											
		$0	$1	$2	$3	$4	$5	$6	$7	$8	$9	$10	
P2 Chooses	*ACCEPT*	$10	$9	$8	$7	$6	$5	$4	$3	$2	$1	$0	*P1 profit*
		$0	$1	$2	$3	$4	$5	$6	$7	$8	$9	$10	*P2 profit*
	REJECT	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0	*P1 profit*
		$0	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0	*P2 profit*

Table 4.1: *Payoff chart for a standard ultimatum game with a $10 stake.*

first mover should take in each circumstance, and, correspondingly, what the utility-maximising response of the second mover should be. The equilibrium prediction, then, is the outcome in which each player's utility is maximised relative to other possible choices.

As articulated by Bolton & Zwick (1995), the game-theoretic analysis of ultimatum games begins with three assumptions:

P1. If players are driven by utility-maximisation, and more money means more utility, then both players always prefer more money to less.[8]

P2. Proposers know *P1*.

P3. Proposers can calculate the optimal offer.

The equilibrium is therefore obtained as follows. Since both players prefer more money to less (*P1*), the proposer, who knows this (*P2*), should offer the smallest possible amount to the responder, allocating the remaining balance to himself (*P3*). The responder should then accept this allocation, since rejection is inconsistent with the desire to maximise monetary reward (*P1* again) (Bolton & Zwick, 1995). According to the canonical model, this should be the only outcome, and no rejections should ever occur.

One benefit of game-theoretic analysis is that it offers a very precise benchmark for assessing behaviour in the ultimatum game. Yet the key finding from Guth *et al*.'s original study was that players did not conform to the equilibrium; instead, players often chose to offer 50% of the resource to their anonymous counterparts, and, in cases where the monetary distribution was not equal, responders sometimes preferred to reject the allocation. In other words, rather than opting for the monetary-maximising prediction of game theory, participants' decisions indicated a preference for notionally fair outcomes. These findings set the stage for over 30 years' worth of research on ultimatum games, much of which has shown that human behaviour repeatedly violates the canonical economic model in a number of important and eye-opening ways. The results from the literature, which amounts to hundreds of experiments, can be summarised as follows:

[8]The idea that players always prefer more money over less, and will continue to do so irrespective of how much money they already possess, is known as the non-satiation postulate (Smith, 1976).

- Studies consistently report offers between 30–40% of the prize pool (Camerer, 2003), with one meta-analysis of 75 experiments revealing an average offer of 40% of the commodity (Oosterbeek *et al*., 2004).

- An even split is the modal outcome in many studies (e.g. Güth *et al*., 1982; Roth *et al*., 1991; Forsythe *et al*., 1994; Hoffman *et al*., 1994; Croson, 1996; Hoffman *et al*., 1996a; Kagel *et al*., 1996; Larrick & Blount, 1997; Slonim & Roth, 1998).

- Responders frequently reject positive amounts, especially those less than 20% of the total money available (Camerer, 2003).

- Tiny amounts, i.e. equilibrium offers, are almost always rejected (Camerer, 2003).

Taken together, the findings show that the behaviour of human players in controlled experimental settings repeatedly deviates from the predictions of the canonical economic model—offers are larger than predicted and not all offers are accepted. The results can, therefore, be interpreted as showing that individuals care about fairness and reciprocity in addition to their own payoffs; are willing to incur personal cost to meet these desires; and are willing to punish those who do not behave in a suitably cooperative manner (Fehr & Schmidt, 1999; Henrich *et al*., 2001). These general findings have been replicated across a wide variety of cultures (Roth *et al*., 1991; Buchan *et al*., 2004; Oosterbeek *et al*., 2004; Chuah *et al*., 2007; Chen & Tang, 2009) including more than 10 pre-technological and tribal societies (Henrich *et al*., 2001, 2004), suggesting that fairness norms are prevalent throughout human societies. Behaviour also fails to meet the canonical model over a wide variety of experimental treatments (Camerer, 2003). There is, however, some considerable variation in behaviour according to the influence of certain variables—such findings are reviewed in section 4.3 of this chapter.

The UG and its results are important for several reasons. First, the results raise questions about the veracity of the canonical economic model of behaviour. The very fact that human players care about fairness norms is evidence against the assumptions of rational, profit-seeking, and individualistic interests. (As in the classic *homo economicus* model of humanity). Second, since players do not always act in pursuit of economic gain, the UG helps to demonstrate the importance of nonmonetary components of utility in driving behaviour away from the predictions of the standard economic model (Andersen *et al*., 2011). At a more general level, the results align with broader contentions about fairness preferences and intolerance of perceived injustice during negotiations.

Perhaps more importantly for the present thesis, the UG has become something of a workhorse for exploring fairness, both in economics and in the wider social sciences. This is because the effects of a given variable can be benchmarked against both the canonical model *and* the modal outcome of an even split. Variables can thus be gauged by the extent to which they cause behaviour to shift between the extremes of pure selfishness and straightforward equality. This in turn makes the UG an extremely useful tool for examining the myriad factors that impact bargaining *per se*. The basic UG is of further importance because it provides a foundation for analysing more complicated types of bargaining (Forsythe *et al*., 1994;

Croson, 1996). That is, by progressing from one-off to repeated games, one can build an increasingly realistic picture of real-world negotiation. Such an approach has allowed researchers to examine the impact of more complicated phenomena, including learning and reputation effects (e.g. Roth *et al.*, 1991; Knez & Camerer, 1995; Slonim & Roth, 1998; List & Cherry, 2000).

In summary, the UG provides a succinct experimental window on negotiation and has served as an important tool for exploring fairness in pecuniary bargaining experiments. However, very little work has investigated division of other resources in ultimatum-style settings. We propose to study division of labour using an adapted version of the UG. The following subsection presents our novel transfiguration of the UG and considers some initial suppositions regarding how players might behave in this game.

4.2.3 A Division of Labour Ultimatum Game

Taking the structure of the UG as an initial framework for studying dyadic bargaining, we present a division of labour ultimatum game (DLUG), where, rather than bargaining over a commodity that both players are assumed to want to keep (i.e., money), players instead bargain over a resource that we initially assume players may prefer to forgo or reduce (i.e., work).

In the DLUG, two players each begin with a task that involves completing a number of work items independently. Both players have identical tasks and must complete them all in order to fulfil the goals of their assignment and collect a fixed monetary reward. The players are first given the opportunity to collaborate, such that the tasks they complete contribute jointly towards the overall work assignment. This collaboration occurs without affecting their payment, i.e. pay is fixed regardless of how much work each individual contributes. However, before this collaboration can occur, players must agree a division of labour. As in the UG, one player acts as the proposer, who begins by suggesting a division of the workload. The second player becomes the responder, who either accepts or rejects the proposal. In the event of acceptance, players can collaborate and completed items contribute jointly to a shared project. However, should the responder reject the allocation, findings cannot be shared and each player must complete his or her task alone. Either of these outcomes ends the DLUG. Players then move on to complete the work according to the outcome of the game.

Table 4.2 displays the range of payoffs in a DLUG with 10 work items, used here as a straightforward mapping of the typical \$10 stake used in the classic UG. In order to offer an initial analysis of our game, we begin by making the following three assumptions. First, and in the spirit of the economic literature, we assume that players are economically rational in that they wish to maximise their rate of pay per unit of work. Since an individual's payment is not contingent on the quantity of work he or she agrees to complete, players should seek to *minimise* their individual workloads. Our second assumption is that players always prefer less work to more, irrespective of what the work actually involves. Our final assumption is that players are selfish and have no regard for the quantity of work completed by their anonymous counterpart, i.e. the relative size of individual payoffs does not matter. Each person is simply interested in minimising

		P1 Chooses *WORKLOAD ALLOCATION*											
		10	9	8	7	6	5	4	3	2	1	0	
P2 Chooses	*ACCEPT*	0	1	2	3	4	5	6	7	8	9	10	*P1 workload*
		10	9	8	7	6	5	4	3	2	1	0	*P2 workload*
	REJECT	10	10	10	10	10	10	10	10	10	10	10	*P1 workload*
		10	10	10	10	10	10	10	10	10	10	10	*P2 workload*

Table 4.2: *Payoff chart for a standard division of labour ultimatum game (DLUG) involving 10 work items.*

his or her workload to gain the monetary reward.[9]

Following the example of the UG, the equilibrium prediction for our DLUG is derived as follows. Since both players are assumed to prefer fewer work items, a self-interested proposer faced with a total surplus of 10 work items should delegate the *maximum* possible amount (i.e., 9) to the responder.[10] As this offer confers an initial 10% reduction in workload, any responder faced by such an offer should accept, as rejection would be inconsistent with the aim of minimising work. As with the UG, this should be the only outcome, and no rejections should ever occur. Players then proceed to complete their allocations, with each person receiving their reward after completion of their agreed individual workload.

The way in which players might behave in the DLUG is currently an open question. On the basis of findings from the classic UG, we might expect that the equilibrium outcome identified above would be relatively rare, with offers instead converging around the point of an even split. This is because, as we saw earlier, responders in the UG typically reject tiny offers due to their apparent unfairness, and proposers, who usually anticipate this behaviour, try to avoid rejection by proposing equitable or near-equitable splits. It is also worth noting that there are several differences between our model and the classic UG, and it would be interesting to explore the impact of these changes. For instance, the UG is strictly a measure of distributive fairness, and there is no further interaction once the responder chooses their action. Yet players in our game must actually proceed to complete the work in accordance with their agreed outcome. Perhaps this additional stage might provide insights regarding distributive *and* procedural fairness, beyond what is capable with the standard UG. Additionally, players would need to engage in some basic collaborative interactions to manage their division of labour, perhaps to prevent redundancy if the task demands it. The need to communicate while enacting allocations might also be important, in terms of ensuring agreed

[9] We regard these as simplifying assumptions, and we do not wish to make an extended argument for work minimisation to be the *de facto* approach to the DLUG. Nor do we argue that this is the correct or most socially acceptable way for players to behave when dividing labour. Rather, we aim to specify how a rational DLUG player might behave if he or she is solely interested in work minimisation (or, by extension, optimisation of wage), in much the same way that the canonical economic model is used as a benchmark for traditional economic experiments.

[10] Recall that the minimum offer in a UG is set to $1 so as to provide the responder with an incentive to accept. We adopt a similar approach to the DLUG, with a maximum work allocation of 9 items. If the maximum offer were all 10 work items, there is no clear reason for a responder to accept beyond pure altruism.

allocations are kept or by allowing said agreements to be renegotiated.

In the following chapter, we put these considerations to the test by implementing and exploring our DLUG in a series of empirical studies. The remainder of this chapter lays the foundations for the studies in three ways. Because the UG is known to be sensitive to many aspects of experimental design, the following section provides a detailed review of factors that are known to impact allocations in UGs—such a review allows for proper control of these issues in our own experiments. We then consider some of the primary theoretical explanations for UG behaviour. This allows for an informed discussion when appraising our results later in the thesis. We then reflect on the merit, and prior use of, experimental games in HCI research. The chapter closes by considering how our model elaborates on this prior work.

4.3 Factors Affecting Behaviour in Ultimatum Games

This section delves into the literature on ultimatum games to provide a clear synopsis of the available evidence. We offer this review as it exposes factors relevant to our game and demonstrates that we did not utilise the game in ignorance of prior work. The review allows us to understand the impact of a wide range of independent variables—an immediate implication is that such variables will need to be controlled if we are to adapt the game successfully. Thus, after describing empirical results associated with each variable, we will sketch a brief methodological implication to illustrate how the findings will inform the experiments reported in the following chapter. Subsection 4.3.5, found at the end of the review, then considers the theoretical implications of the surveyed results in terms of their relevance for the present thesis.

In addition, the body of work we will describe further illustrates how the prediction of the canonical economic model, which views 'rationality' as strictly profit-seeking, is almost never supported. Instead, what will become apparent is that people appear to behave 'rationally' in a social, rather than economic, sense. For example, the economic model predicts that people should accept any offer, yet small offers tend to be rejected in practice. While rejection of a small offer is costly in economic terms, it is likely to be socially rewarding because it presents an opportunity to give payback for an insulting offer. It is therefore not surprising that people are content to punish small offers because the social payoff associated with rejection is likely to outweigh the financial gain that would stem from accepting. However, if the right to make small offers was *legitimised* through an established and mutually recognisable claim to a greater share of the resource, it might be considered socially *irrational* to respond with rejection because such a decision would ignore one individual's entitlements (cf. our definition of fairness from Chapter 2). The ability of the UG to tease out such complexities is something that is very much demonstrated within this review. We will examine the ways in which different variables can cause allocations to shift away from (and towards) the modal outcome of an even split, and we will see how superficially 'unfair' offers, as characterised by inequity, can be legitimised as fair according to various influential factors. Exploration of these issues is not only illustrative in terms of theory but also provides a backdrop to the experimental variables explored in Chapter 5, allowing for considered interpretation of our own findings.

The present review is organised in accordance with three categories of variable: *methodological and* *structural*, *descriptive*, and *demographic*.[11] We also introduce the dictator game (DG) (cf. Kahneman *et al*., 1986a), the structure of which is described in more detail in subsection 4.3.1 below. In brief, the DG is a simple variant of the UG in which responders cannot reject the allocation. In this way, it presents a very clean test of distributive generosity. We feel it is appropriate to consider the DG in this review because of its close relation to the UG (many studies employ both to compare behaviour across different treatments) and because the DG serves as inspiration for one of the experiments reported in Chapter 5 of this thesis. While the review is scoped to consider the ultimatum and dictator games, it is worth noting that the general findings are applicable to other economic games, e.g. the aforementioned public goods and prisoner's dilemma problems.

4.3.1 Probing Fear of Rejection: The Dictator Game

The dictator game (DG) (cf. Kahneman *et al*., 1986a) is a simple variant of the UG in which the responder's opportunity to reject an offer is removed from the bargaining situation. Thus, the proposer in a DG simply chooses how to divide the surplus and the responder, who has no recourse, receives whatever has been allocated. The equilibrium in a one-shot situation with a $10 stake is easy to derive: the proposer keeps everything and the responder receives nothing whatsoever.[12] This model has been used to explore the extent to which the high number of equitable offers in the UG can be explained by proposers' fear of rejection (e.g. Guth & Tietz, 1988; Ochs & Roth, 1989; Kravitz & Gunto, 1992). Fear of rejection is of interest to economists because, if true, it would suggest that fairness in the UG arises from strategic thinking, rather than as a result of prosocial preferences for equity and sharing (Hoffman *et al*., 1996b).

The first comprehensive comparison of ultimatum and dictator game giving was effected by Forsythe *et al*. (1994). In their experiments, 65% of UG proposals resulted in even splits, compared to just 22% in the DG. The percentage of heavily inequitable offers was also much higher in the DG: overall, 36% of dictators gave nothing to their counterparts, with a further 30% offering a tiny fraction of the pie. Such offers were entirely absent from the UG treatments.

Hundreds of dictator experiments have been reported since the original work of Forsythe *et al*., and the results continue to support the original hypothesis about fear of rejection. For example, a recent meta-analysis of 616 DG treatments found that the average offer was 28.35% of the pie (Engel, 2011). When this figure is compared to the mean 40% offer in ultimatum games (Camerer, 2003; Oosterbeek *et al*., 2004), it seems clear that people are less generous on average in dictator games. However, it is worth noting that the results are still well above the offer predicted by the canonical economic model. This implies that, although the fear of rejection can partly explain the frequency of equitable offers in the UG,

[11] We wish to acknowledge that the structure of this section is loosely based on a prior review of economic literature by Nick Wilkinson (2008). However, our review differs to his work in that our focus is solely on the ultimatum game and its foremost variant, the dictator game, and we omit certain variables that cannot be explored using these games. Our review is also more up-to-date and incorporates literature from 2008 onwards.

[12] Note that this differs to the UG in that a dictator game typically has no minimum offer. The UG requires a minimum offer of $1 because, without an incentive to accept, the game simply becomes a test of the responder's taste for punishment. A $0 allocation in the dictator game, however, is meaningful because it demonstrates total selfishness on the part of the dictator.

other factors are at work and people do still behave somewhat altruistically even when they cannot be punished for selfishness.

For our purposes, the DG is important because it is akin to a control condition for the UG: by comparing results from the two, experimenters can "observe how much more people propose because of fear... as opposed to, loosely, simply being generous" (Kurzban, 2013, online). Because the differences between ultimatum and dictator games are interesting from a theoretical standpoint, and because a majority of experiments use both games, the remainder of this review shall consider the extent to which our surveyed variables impact behaviour in both of these games.

4.3.2 Methodological and Structural Variables

Variables in this category change how experiments are conducted and are important since they are for the most part controllable. This means that, through proper manipulation and experimental control, individual factors and their effects can be isolated and studied systematically.

4.3.2.1 Repetition and Learning

A large majority of ultimatum and dictator experiments are typically run as one-shot affairs, where participants arrive at the lab, play one round of the game, and then leave with their respective earnings. Several authors (e.g. Roth *et al.*, 1991; List & Cherry, 2000) have argued that such experiments do not provide ample opportunities for learning; the supposition is that, if participants were afforded more time to shed their naivety and learn about the game through repeated trials, behaviour might converge towards the equilibrium prediction of the canonical model.

The effect of learning can be studied in two ways: either by iterating resolved games to examine inter-game learning, or by allowing proposers to iterate their offer in a single game, thereby studying intra-game learning effects. Regarding the former, studies have found a slight tendency for offers and rejections to fall over time, suggesting convergence towards self-interest after repeated play. For example, Roth *et al.* (1991) observed a small decrease in offer size from players from four different cultures playing ultimatum games over 10 rounds. In a different experiment, Slonim & Roth (1998) found that, over the course of 10 rounds of ultimatum play, offers declined in tandem with increased experience but only in high stakes games for a week's wages. A similar result was obtained in another 10 round experiment by List & Cherry (2000), who observed that responders became more willing to accept inequitable offers over time. However, tiny offers were still rejected about 60% of the time, with players only willing to accept offers in the region of 20–30% in later rounds.

Contrasting evidence from other authors suggests that players may in fact become *more* generous if given the opportunity to learn. For example, Oosterbeek *et al.* (2004) found that when UG results were aggregated and meta-analysed, repeated play actually *increased* the proposed share and did not significantly impact rejection rates—this directly contradicts what the learning argument would predict! More recently, Brenner & Vriend (2006) conducted an experiment where groups of proposers played 100 rounds of the

UG against computer-generated agents. The humans knew they were playing against computers—this, according to the researchers, should remove all considerations about fairness and reciprocity.[13] Results from four groups of human players showed that players did not learn to make the equilibrium offer, and were instead boundedly rational, testing the problem space by making a variety of offers and then persisting with the ones that provided 'good enough' payoffs without provoking rejection.

Turning to intra-game repetition, Gneezy *et al*. (2003) studied a 'reverse' ultimatum game in which proposers had multiple chances to offer an allocation to responders. The game followed the structure of a standard UG, except that, when a responder chose to reject, the proposer was able to make further offers until either the responder accepted or the proposer decided to terminate the game (in which case neither party received anything). However, the proposer was only permitted to increase his or her offer after a rejection, meaning that responders should reject offers until the maximum possible amount is allocated. As it transpired, when participants bargained over a surplus of 25 tokens, the modal accepted offer was an allocation of 13 chips, with 12 the next highest (i.e., the two possible splits closest to equality). When participants were given a deadline to encourage last-minute, delayed offers, average allocations were higher than expected: 11.5 and 10.4 chips under three- and one-minute deadlines, respectively. The study does not support the learning argument since responders did not continually reject offers in the baseline condition, and, with deadlines, proposers did not wait until the last minute to make the utility-maximising offer.

In sum, the evidence in this area is mixed, meaning that the impact of learning is unresolved—this issue is a current research area in experimental economics (e.g. Chen & Gazzale, 2004; Mengel, 2012; Grimm & Mengel, 2012). What is clear is that, even in studies where learning effects were found, the extreme predictions of the canonical model were unsupported. While repeated play is important, one-shot games are a better starting point because they allow for control of punishment strategies (Roth *et al*., 1991) and reputation effects (Hoffman *et al*., 1996a). For our purposes, this means that it will be best to begin with a one-shot version of our DLUG before attempting iteration.

4.3.2.2 Anonymity

In economic games, the behaviour of participants may be influenced by a lack of anonymity. The knowledge that decisions are under scrutiny could cause participants to behave in a certain way so as to appear 'nice'. Similarly, participants may seek to understand the object of the study in order to appease the experimenter or maximise their chances of receiving invitations to future studies (cf. Hoffman *et al*., 1994). These concerns are especially problematic for economists seeking to draw conclusions about underlying preferences—fair offers may not reflect participants' preferences but may instead represent attempts at playing up to what they believe is the 'correct' or most socially desirable behaviour (Levitt & List, 2007).

Some fairly elaborate experimental designs have been created in order to test for the effects of

[13] It is worth noting that people have a tendency to treat computers as if they were people (Reeves & Nass, 1996), and it is not, therefore, necessarily certain that all 'emotional' concerns regarding fairness were removed.

anonymity. While economics experiments usually involve a reasonable level of anonymity anyway, multiple studies have sought to explore the impact of perceived scrutiny by tightening the level of anonymity between participants and the experimenter. In an early example, Hoffman *et al*. (1994) used a dictator game involving double-blind anonymity, where participants were required to place their allocations into envelopes and then insert the envelopes into a sealed box. This process guaranteed complete isolation of individual decisions—no individual, including the experimenter or any later observer of the data, could know who had made which decision, and participants were aware of this fact. The researchers found that players were much more selfish in the double-blind condition: 62% of the allocations were made at the equilibrium ($0), compared to a control group with just 8%. Bolton & Zwick (1995) report similar results for ultimatum games—double-blind anonymity resulted in a 10% increase in zero offers, as compared to a control condition. However, a more recent meta-analysis of dictator games by Engel (2011) did not find a significant effect for double-blind anonymity on benevolence unless the experiment was iterated. Thus, while scrutiny from the experimenter does appear to result in more socially acceptable allocation behaviour (i.e. fairness), it is possible that anonymity only matters in more complicated designs.

Other work has shown how relaxing the level of anonymity between participants can influence fairness. Bohnet & Frey (1999) designed a study with three treatments: a standard anonymous DG; a DG where the recipient was identifiable to the dictator; and a two-way identification where both dictators and recipients could identify one another visually. The researchers also varied the procedure in the one-way identifiability condition, whereby some recipients held a number in their hands so that dictators could identify them. Others also stated their name, city of origin and academic major. In the high-anonymity condition, 28% of dictators kept all the money for themselves, whereas none of the dictators in the one- and two-way identifiability cells chose to keep all of the money. Participants in the two-way identifiability condition were most generous, with over 70% giving away half of their resource. Other studies have found that generosity increases if participants feel as though they are being watched (Haley & Fessler, 2005), are knowingly filmed (Wilson *et al*., 2010), or are primed with material about omniscient dieties (Shariff & Norenzayan, 2007) while making their decisions.

Based on these results, it appears that anonymity does affect allocations, although behaviour still does not conform to the canonical model. Participants are more likely to behave in line with fairness norms when they feel as though their behaviour is scrutinised, but when decisions are perceived as anonymous and unidentifiable, participants are more likely to become selfish. However, one problem with the studies in this area is that it is unclear whether these behaviours stem from greater distance between participants and the experimenter, or from participants and other participants. In other words, the conceptual differences between whether participants are concerned about being identified by the experimenter, by other participants in the experiment (who may or may not be peers), or indeed by both, are quite poorly defined. The only way to navigate this issue would be to manipulate these two aspects independently—one might expect that decisions would be highly selfish when participants feel as though they are entirely anonymous and free from scrutiny, though no work has yet attempted this.

In terms of our own work, the present results emphasize the need to ensure that anonymity is controlled between, and within, conditions. We identify two courses of action: first, participants in our experiments should have anonymity from one another—this will allow for the impact of identifiability to be assessed at a later date. Second, anonymity between the participants and the experimenter needs to be fixed. We opt for two-way identifiability between subjects and facilitators. While this may increase feelings of scrutiny among participants to levels beyond what they would experience in everyday life, it alleviates the need for complicated double-blind procedures. We can also suppress this scrutiny by indicating in the experimental protocols that participants should keep their decisions private and not read them aloud to the facilitator.

4.3.2.3 Communication

Prior work has identified that communication has a strong impact on benevolence in experimental games (Sally, 1995), and it is for this reason that participants are not usually allowed to communicate beyond the written exchange of their offers. In both the UG and the DG, communication typically leads to an increase in the number of equitable offers. In a study by Bohnet & Frey (1999), dictators listened to recipients talk about themselves and their academic major—this manipulation led to a fair split becoming the modal outcome, with 40% of dictators actually giving *more* than half. However, an earlier study by Frey & Bohnet (1997) found that when dictators were faced with two recipients, generosity was shown only to the recipient with whom dictators were allowed to communicate. This clarifies that communication results in target-specific sympathy rather than a general feeling of generosity (Wilkinson, 2008).

Other work has explored the impact of 'cheap talk' (lies, threats, and promises) in UGs. Croson *et al*. (2003) used a four-round UG with varying pie sizes and outside options for responders (an outside option is a sum, typically 10–20% of the surplus on offer, that is guaranteed to the responder should they choose to reject an allocation). Both sets of players were allowed to pass notes to their respective partners in order to exchange information about their intentions. There were three findings: first, when responders lied about the size of their outside option, offers from proposers were larger. Second, threats of rejection from responders significantly increased offers. Third, proposers who lied about the size of the pie were able to make significantly lower offers without lowering the probability of their acceptance. These results are perhaps best regarded as demonstrating the mediating effects of communication on cooperation; while it can be used to deceive and gain unfair advantages, it can also be used to enforce compliance with social norms, as with those responders who threatened to reject unsatisfactory offers.

In sum, the impact of communication is such that experimenters try to minimise it by, as mentioned above, disallowing all communication beyond the written exchange of offers and acceptances among participants. The implication for our work might be that participants should not be allowed to communicate. However, this would be very restrictive in terms of exploring collaborative work, which, as articulated in Chapter 2, is underpinned by communication. We will strive for a faithful replication of the procedure used in economic methods by restricting communication during the DLUG allocation procedure. Participants will then be allowed to communicate freely during enactment of the agreed workloads. This will offer a sufficiently rich picture of collaboration while avoiding the potential impact of friendly communication on the DLUG outcome.

4.3.2.4 Stake Size

Recall that the typical financial stake in a UG experiment lies in the region of $10. Sceptics of economic experiments argue that such a relatively inconsequential sum is incapable of inducing realistic behavioural responses because the amount to be rejected is a pittance (see Camerer & Thaler, 1995, page 210). The supposition, then, is that bargaining over high stakes will create a shift towards self-interest, and hence the predictions of the canonical model. Similarly, responders should also be less likely to reject if the cost amounts to more than a couple of dollars.

The evidence from studies exploring this question shows only a weak effect for stake size, consistent over a variety of pecuniary amounts and research populations. An early test by Forsythe *et al*. (1994) found that the size of the pie ($5 versus $10) did not affect distribution of offers. Similarly, Hoffman *et al*. (1996a) found that allocations were comparable between low and high stake sizes ($10 versus $100) and not all offers were accepted in the high stakes condition. In some cases substantial offers were rejected, including two out of five $30 offers in the high stakes condition. Overall, however, when the data were combined with that from two other studies (Forsythe *et al*., 1994; Hoffman *et al*., 1994), the only significant result was that responders were less likely to reject when stakes were higher. More recent meta-analyses confirm these findings for both ultimatum (Oosterbeek *et al*., 2004) and dictator games (Engel, 2011).

However, the aforementioned studies are problematic because rejection of $30 is not that high compared to, say, rejecting hundreds or even thousands of dollars. Other studies have explored this issue by decamping to countries where the typical stake used in an American experiment has a much higher purchasing power. Some studies suggest that increasing stake size does not impact behaviour: a study conducted in the Slovak Republic (now Slovakia) by Slonim & Roth (1998) compared ultimatum play over 60 Slovak crowns (SK, equivalent to 10$ at the time of the experiments) and 1500SK (equivalent to the average weekly wage for Slovaks at the time). They found no significant difference between offers in low and high stakes games during one-round games, though responders were about 10% less likely to reject in high stakes conditions. A similar experiment in Indonesia by Cameron (1999) raised the stakes to three times the monthly expenditure of an average participant. Overall, Cameron's results were similar to the Slonim & Roth study, in that a fair offer was the modal outcome and no significant impact was found for stake size on proposer behaviour. Responders were also more willing to accept unfair offers in the high stakes condition.

Alternative evidence comes from List & Cherry (2000), who examined the dual impacts of stakes and learning. They found that, over the course of 10 rounds of ultimatum play, proposers did become more selfish with their offers. However, this effect only occurred when the stake was $400—self-interested behaviour was much less common with a $20 stake. They also found that, although stakes had no impact on rejections of tiny offers (such offers were almost always rejected), the mean rejection rate decreased in the high stakes condition, indicating that responders were more willing to accept lower offers when the stakes were higher. List & Cherry explain their results using cognitive cost theory, intimating that the cost of deriving the equilibrium strategy is too high when stakes are low. It is possible that the increase in stakes makes the decision seem more important—prior work does argue that heuristics are less likely to be applied in important decisions than in trivial or unimportant ones (Sherman & Corty, 1984). However, it is still interesting that the behaviour of participants in List & Cherry's study was tempered by fairness concerns—people did not become totally self-interested. A more recent experiment by Andersen *et al*. (2011) also found that high stakes leads to inequity. They conducted an experiment in several villages in Meghalaya, Northeast India, with a maximum stake size of 20,000 rupees, equivalent to just over a year's income in the villages in question. They found that offers were significantly lower with high stakes when compared to other low stakes environments, and there were significantly fewer rejections in the year's wage condition.

Based on these studies, the canonical model is not supported unless the stake is very high and stands to have a major impact on an individual's life. For responders, the pecuniary reward may dominate the punishment value at higher stakes, but may be outweighed by the punishment value at lower stakes. However, stakes may not have this effect "if, as stakes increase, a responder's utility from punishing a proportionally small offer rises at least as much as his utility increases" (pp. 578 Slonim & Roth, 1998). Fairness perceptions, then, still matter at high stakes, but people are certainly more willing to tolerate inequity if the amount they each receive is sufficiently large.

Although the evidence in this area is mixed, the studies detailed above at least imply that stakes are worth controlling. It is not difficult to imagine that participants could become significantly more likely to delegate tasks if their workload is unreasonably high or if the work itself is especially arduous. As we do not yet know the exact impact substituting money with work, we will use a straightforward mapping of 10 work items in place of 10 units of currency. We can only speculate as to the relative utility of contributing each work item to a joint project, but using a 10 item task will at least allow for superficial comparisons between our results and the $10 stake used in most UG experiments.

4.3.2.5 Entitlement

As outlined in Chapter 2 of this thesis, fairness is based on equality but also accounts for the entitlements and needs of each involved party. Researchers have explored the impact of perceived entitlements, both to roles and resources, on ultimatum bargaining. Based on our own understanding of fairness, we might expect that those who perceive themselves as bearing a legitimate claim would attempt to retain more of a surplus when asked to divide it. Empirical evidence suggests this to be true, and a variety of techniques have been employed to induce the feeling of entitlement.

Güth & Tietz (1986) (cited in Hoffman *et al*., 1994) used a design where the roles of proposer and responder were independently auctioned to the highest bidder. They found that participants generally paid more for the right to be the proposer, and the recognition that proposers had paid a higher price led to smaller offers and increased rates of acceptance. An alternate approach was employed by Hoffman *et al*.

(1994), who asked participants to complete a general knowledge quiz prior to playing the UG. Participants with the highest quiz scores were awarded the role of proposer, and the researchers expected that this sense of deservedness would lead to smaller offers. On average, offers fell by about 10% when the proposer role was earned, as compared to a control condition where the modal outcome was a fair split. Moreover, this decrease in offer size was achieved without significantly raising the frequency of rejections, implying that the entitlement gained perceived legitimacy in the eyes of both participants.

Evidence from studies of the dictator game provides similar insight regarding entitlements. An experiment by Cherry (2001) used two conditions, one a standard DG, and the other a DG where dictators completed an earnings task which involved making bets on the outcomes of a lottery. Total earnings were then based on the results of the betting. In the baseline condition, 74% of dictators allocated a positive amount, with 14% offering half of the money. In the earnings treatment, however, only 24% of dictators made a positive offer and there were no equal splits. Similar results were obtained by Cherry, Frykblom, & Shogren (2002) who found that the number of zero offers rose from 17% to 80% when dictators earned their endowments, increasing to 96% when the earnings task was combined with a double-blind anonymity procedure. Evidence from a recent experiment (Carlsson *et al*., 2013) suggests that behavioural differences concerning earned wealth remain true in the field; dictators were significantly less likely to give their payment away when it was earned.

Other work has explored perceived entitlements claims for both dictators *and* recipients. Oxoby & Spraggon (2006) found that when dictators earned the right to claim the wealth by completing a quiz (but recipients did not) dictators kept all of their money 100% of the time. But, when the recipient's claim to the money was established via an earnings task (which was not completed by the dictator in this condition), just 5% of dictators kept all the money, and, in some cases, dictators gave away all of the endowment to the recipient. List & Cherry (2008) report a similar experiment except both participants completed an earnings task. They found that approximately 90% of dictators subsequently gave nothing, as compared to 70% in a condition where only dictators earned their wealth and 50% in a baseline condition. One other study by Eckel & Grossman (1996) revealed that offers increased if a responder is perceived as especially deserving. When dictators were offered the opportunity to allocate some of their \$10 endowment to an anonymous student, participants donated an average of 10% of their payoff, but when the recipient was identified as the American Red Cross, participants donated an average of just over 30%. While the fact that allocations increase if a benefactor seems more deserving of aid feels intuitively obvious, the donation of more to a charity is theoretically interesting as it suggests that the moral cost associated with *not* giving money to a charity may be higher, meaning that aversion to feelings of guilt may guide offer sizes.

The findings concerning entitlements are one of the more reliable manipulations to be found in the literature, and because entitlements can shift allocations both towards and away from equity, the results are again problematic for the canonical model. We argue that entitlement manipulations actually demonstrate compliance with fairness norms; equity is not always the fairest outcome, especially when entitlements are present. In our own experiments, we will not manipulate perceived entitlements in our baseline conditions. This will provide a clean test of participants' preferences, as revealed in the absence of rationale that would suggest one allocation over another.

4.3.2.6 Number of Players

Several studies have incorporated multiple responders to explore how players evaluate their own payoffs relative to those of others. For example, Knez & Camerer (1995) used a design involving a single proposer who made simultaneous offers to two responders. The researchers found that responders disliked relative inequity and would typically compare their own payoffs; if the difference was too high, rejection was significantly more likely. Even a 50 cent increase in the first responder's payoff relative to the second's doubled the likelihood of rejection. However, another finding was that responders were even more concerned about their own payoff relative the proposer's, and another study in three-person games involving an active and non-active responder showed that allocations of between 10–15% to non-active recipients were only rejected by active responders about 5% of the time (Güth & Van Damme, 1998). This indicates that players tend to be mainly concerned about fairness in relation to active others, rather than to all involved parties *per se*. In the present thesis, our experiments will focus on two player bargaining to restrict the impact of having multiple counterparts in the work scenario.

4.3.2.7 Available Information

The information held by each player can also impact allocations. In ultimatum games, information is usually explored by introducing an *asymmetry* where one player knows more than the other about some aspect of the game. Regarding the amount of money to be divided, responders may be in one of three states of knowledge: perfect information, where the amount on offer is known for certain; imperfect information, where the possible pie sizes and their probability distribution are known; or no information, where no knowledge about the pie size is held whatsoever. In the latter circumstance, one might expect that proposers could lower their offers without risking the threat of rejection. Responders might also become more willing to accept low offers as they have no way of comparing their payoff to that of the proposer.

The general finding concerning this manipulation is that responders are more likely to accept smaller offers under circumstances of incomplete or no information (e.g. Mitzkewitz & Nagel, 1993; Straub & Murnighan, 1995; Croson, 1996; Rapoport *et al*., 1996; Croson *et al*., 2003). Since small offers could indicate a small pie size, and no opportunity for social comparison exists when the pie size is not known, responders seem willing to give proposers the benefit of the doubt and would rather receive some money than none. Of greater interest is the behaviour of proposers: Croson (1996), for example, observed that offers were significantly lower when proposers knew that responders could not determine the total size of the pie. A subsequent study showed that this behaviour was also sensitive to knowledge about a responder's outside option, with proposers offering about 10% more when the exact size of the option was known (Croson *et al*., 2003).

Information can also be manipulated by concealing or varying participants' knowledge about the relative value of payoff items. Kagel *et al*. (1996) investigated fairness by using a design where UG players were tasked with dividing 100 tokens. The experimenters manipulated both the value of the tokens (a single token was worth either 10 or 30 cents to each player) and the information given to each player about relative values (i.e. whether or not both players are aware of the true value of tokens in their respective trial, and whether or not this awareness is common knowledge). When both players are paid equally for their tokens, and both are aware of this fact, the situation becomes a standard UG, with 100 tokens equivalent to $10. But things are quite different when each player's chips are of a different value. If a proposer knows that chips are worth 30 cents to him but only 10 cents to a responder, a truly fair offer would require an allocation of 75% of the chips. Such an allocation would mean that the monetary payoff each receives from the tokens finishes up equal. But since only the proposer knows the true value of his tokens, he can offer exactly half and *appear* fair while actually receiving more money than the responder. This outcome is exactly what was observed during the study—the mean offer in the asymmetric condition was 46%, meaning that the proposer's take home pay is actually much greater than the responder's. Since responders had no reason to believe this was an unfair offer, rejections were almost non-existent. This study supports the argument that proposers merely want to seem, rather than be, fair, and is important as it demonstrates how information asymmetry can sometimes lead to self-interested behaviour.

Finally, information can also be provided about the past histories of players, in terms of their previous offers or behaviour. Such information can provide a reference point for fairness judgements, and brings bargains closer to real-world situations where reputation effects are likely to play a role. Perhaps not surpisingly, several studies have found that players tend to offer more to other players who have a visible history of fair play (Knez & Camerer, 1995; Cason & Mui, 1998) and in the study by Croson *et al*. (2003) reported above, responders demanded more and were more likely to reject offers from proposers who had lied and behaved unfairly in previous ultimatum bargains.

For our own work, information will be easy to control—we will simply ensure that players have complete knowledge of the payoffs and that the information given to players is equal. This will allow for future manipulations of information availability. (Such a manipulation is explored in Thesis Study 4, reported in the following chapter).

4.3.3 Descriptive Variables

Descriptive variables are those that alter superficial aspects of the experiment, particularly in the way the situation is presented to participants. Here we focus on framing effects induced via changes to the experimental protocols.

4.3.3.1 Framing Effects

Framing effects, which refer to simple changes in the descriptive language used to present an action or a series of options, have been studied extensively in the literature on contingent weighting in judgement and choice (cf. Tversky & Kahneman, 1981). The main finding is that frames can affect the outcomes of choice problems, to the extent that actions can be systematically altered according to the way in which a particular decision outcome is framed (see Tversky *et al*., 1988).

Regarding ultimatum and dictator games, studies have shown that procedural frames, which are defined as "ways of describing actions in structurally equivalent allocation procedures" (Larrick & Blount, 1997, p. 810), can suppress equitable allocations. Hoffman *et al*. (1994) examined how varying the instructions in an ultimatum game can negatively impact equitable offers. Their first condition used standard instructions that asked participants to 'divide' the resource; the second used instructions framing the experiment as an 'exchange' between buyers and sellers, thought to induce more self-interested behaviour by invoking behavioural norms associated with competition. The study showed that the average offer size was significantly lower (by almost 10%) when the game was framed as a buyer/seller exchange, with the percentage of equitable offers falling by about 30% overall. A similar effect was observed in dictator games: about 40% of dictators made equitable splits with standard instructions, yet with the exchange instructions there were no fair offers and zero allocations increased by 20% (Hoffman *et al*., 2008).

Procedural frames can also *enhance* cooperation in games. Larrick & Blount (1997) used two conditions: one with the standard 'division' language, and a second where players 'claimed' an amount of the resource. The researchers found that 56% of splits were equitable in the claim condition compared to 35% in the divide frame. Participants were also more likely to accept zero offers under the claim language. The authors argue that the 'claim' language implies equality of ownership, whereas the divide language is more confrontational—perhaps proposing a 'division' implies that player one has control over player two's outcome.

Finally, a study by Croson (1996) examined the effect of making proposals in percentages versus dollar amounts. Evidence from the contigent weighting literature suggests that when offers are made in percentage terms, making relative payoffs more salient, greater weight is assigned to fairness considerations (Tversky *et al*., 1988; Croson, 1996). Croson observed that the framing of an offer as a percentage caused rejection rates and average demands to increase, as compared to offers framed as pecuniary amounts. These results were consistent with the predictions based on contigent weighting—when fairness is made salient via expressing allocations as a percentage amount, equitable allocations increase.

We will make use of these findings in three ways. First, we will use instructions from prior UG experiments as a basis for own studies. This will ensure that our protocols are couched in the *lingua franca* of the related literature. Second, we will keep language consistent from study to study, i.e. by not arbitrarily changing from 'allocating' work to 'delegating' or 'claiming' workload. Finally, we will use absolute figures rather than percentages when describing participants' workload assignments.

4.3.4 Demographic Variables

Numerous studies have explored the impact of demographic variables on ultimatum and dictator behaviour. Here we explore three variables that may be relevant to the design of our own experiments.

4.3.4.1 Gender

The majority of studies have found no differences in the relative size of offers made by men and women when playing with anonymous partners, in both ultimatum (Eckel & Grossman, 2001) and dictator games (Frey & Bohnet, 1995; Bolton *et al*., 1998; Dufwenberg & Muren, 2006). That said, women tend to reject less often and are more likely to accept low offers (Eckel & Grossman, 2001). When the gender of the responder is known to the proposer, both genders tend to give more to men than women (Eckel & Grossman, 2001; Solnick, 2001; Ben-Ner *et al*., 2004). Schweitzer & Solnick (1999) further found that females offer about 5% more to attractive males, as compared with unattractive males.[14] For dictator games, a recent meta-analysis that did not control for anonymity showed that women both give and receive more on average (Engel, 2011). In our own work, it is likely that the impact of gender will be mitigated by our decision to keep participants anonymous.

4.3.4.2 Education

Some studies suggest that academic major can influence allocation behaviour, and it is certainly possible that, because of their studies, students of economics and business might be more aware of the game-theoretic solution in a bargaining experiment. However, the evidence on this issue is mixed: Carter & Irons (1991) observed that students of economics offered about 7% less and demanded 7% more than noneconomists, whereas Kahneman *et al*. (1986b) found the reverse, with students of economics and business actually more generous than those studying psychology. Other studies have found no differences between economics students and those from other disciplines (e.g. Eckel & Grossman, 1996; Kagel *et al*., 1996; Oosterbeek *et al*., 2004). We will control for education by ensuring that our participants are not drawn from a single academic discipline or background of study.

4.3.4.3 Culture

Several landmark studies suggest that cultural norms of fairness do impact bargaining outcomes in the ultimatum game. The main finding is that there appears to be cross-cultural variation in what constitutes a ‘fair’ offer. Roth *et al*. (1991) examined UG behaviour in four countries: America, Israel, Japan, and Yugoslavia (now Slovenia). Although the mean and modal offers were between 40–50% of the prize pool in all four cultures, a small amount of cultural variation was present in the data; proposers in Israel were more miserly, offering an average of 10% less than participants in the USA and Yugoslavia, and rejections were also less common in Israel. However, one problem with studies of this sort is that they only survey a single city and may not reflect broader intra-cultural differences: Buchan *et al*. (2004) found results contradicting those of Roth *et al*. (1991), e.g. that Japanese were more generous than American participants. More recently, a meta-analysis of 37 UG studies drawn from populations across 26 countries actually found no significant differences in proposer behaviour between cultures, though there were significant differences

[14] This study is unique in that it is the only recorded case in which the modal offer was more than 50% of the pie.

among rejection rates (Oosterbeek *et al*., 2004).

More interesting evidence comes from a comprehensive comparison of 15 small-scale, pre-technological and tribal societies by Henrich *et al*. (2004), who found a number of intriguing results concerning average offers in ultimatum games. Two tribes, the Lamelara of Indonesia and the Ache of Paraguay, were extremely generous and frequently offered more than 50% of the prize pool to responders. Conversely, the Peruvian Machiguenga tribe offered an average of just 26% of the prize. The researchers link these findings to the societal characteristics of the tribes in question, focusing on two explanatory variables. First, the amount of general cooperative activity among citizens: when members of a society hunt in groups, ultimatum offers tend to be higher. The most tight-fisted tribe, the Machiguenga, rarely trade outside the immediate family. Second, the degree of market integration is important: when a large-scale market exists for the trade of goods and labour, ultimatum offers tend to be higher.

Overall, the evidence for culture suggests that perceptions about fairness differ between societies, and that these perceptions directly affect behaviour. While there is mixed evidence for cultural variation in technologically developed societies, it is clear that no culture ever conforms to the predictions of the canonical model. Although it would be easy to control for culture in our experiments, we prefer to invite participants from across our campus as we do not have access to a guaranteed subject pool. Thus we must rely on the persons that we are able to recruit through our own endeavours.

4.3.5 Interpretation & Implications for this Thesis

The previous subsections reviewed a total of thirteen relevant factors that are known to impact behaviour in the UG. Some of these factors are the result of extensions to the basic blueprint, e.g. iterating resolved games or changing the value of the stakes on offer. Others are simply a matter of who is playing the game, e.g. participants' gender or country of origin. While the review has been offered partly as evidence of our having thoroughly surveyed the UG literature, it also allows us to understand what the impact of each variable is and whether or not it is particularly important for our own ends. This feeds into our experimental design, as described more thoroughly in the following chapter.

It is worth considering the implication of results from the UG in terms of what they reveal about human behaviour. Clearly players ascribe to some basic notion of fairness in social exchange, a fact that is made salient by the UG. Control conditions show that offers are frequently made on the basis of equity, and responders happily reject positive amounts to reach an equitable outcome. However, as revealed by the present review, proposers can become more or less likely to make equitable allocations in accordance with particular variables, and responders can become correspondingly more or less likely to reject an allocation. For instance, people will try to increase their own profits when they know that they are not being monitored, or when they can leverage a competitive advantage, e.g. an information asymmetry. But people become more generous if they can see their partner, are allowed to communicate with them, or feel that recipients are especially needy. These findings demonstrate how people selectively invoke different fairness rules according to the situation at hand (cf. Leventhal, 1976). The UG literature helps to pinpoint the circumstances in which particular rules are applied and, in turn, allows one to make inferences about how particular effects would extrapolate to more complicated settings (Camerer, 2003).

The results also expose some of the tensions that exist between individual preferences and the pressures imposed on decision making by societal norms. Such norms are macrosocial structures that exist outside of individual preferences (Binmore, 2010b; Gintis, 2011). Tension occurs between the two because selfish choices often appear preferable but their selection is tempered by norms that suggest otherwise. The UG literature helps us to reason about how fairness might operate with respect to the salience of particular norms. For instance, a feeling of high anonymity likely makes fairness norms for sharing less salient. Knowledge of how a given variable impacts allocations therefore provides ammunition to attack the question of why people behave as they do in situations where different norms may be in operation.

We have already noted that some variables can increase generosity, whereas others can steer allocations towards selfishness and the predictions of the canonical model. In the economic literature, any deviation from equity is sometimes taken as evidence of self-interest, irrespective of the circumstances under which it occurs (e.g. Cherry *et al*., 2002). While it is true that certain manipulations demonstrate a willingness to deceive for selfish gain, the present review has stressed the need to consider *why* apparent self-interest occurs. It is not always the case that superficially ‘selfish’ players are being unfair. Rather, the perception of behaviour as fair or unfair depends on the extent to which a competitive advantage is being leveraged as well as the underlying *intention* behind actions (Rabin, 1993; Nelson, 2002). As noted by Camerer (2003), “Fairness is a judgment people make about an action players take or its consequences, and that judgment affects their preferences for actions and allocations. Whether an action is judged as fair, and what players do as a result, respond to observable variables in intuitive ways. It is fair to keep more if you became the Proposer by winning a contest, or if keeping money is the only way a Proposer can play a second time and earn more money” (p. 114). The latter part of Camerer’s argument runs parallel to proposed rules of justified self-interest, which dictate that inequitable allocations can be fair in the correct circumstances (Lerner, 1971; Leventhal, 1980). The point here is that self-interest is not always ‘selfish’ but may actually be a key component of ensuring fairness norms are upheld—one needs to look beyond the numbers to understand what is motivating different allocation decisions and how those motivations relate to a decision maker’s attempts to operate in line with their understanding of fairness.

Given the wide array of empirical findings, one might wish to turn to an accepted theory that provides a complete explanation of the results. Sadly, such a theory does not exist. This is because the diverse and sometimes conflicting outcomes from the UG have proven difficult to consolidate within a single framework. What attempts there have been in this regard (e.g. Fehr & Schmidt, 1999) still fail to account for all of the empirical data (Binmore, 2010b) and thus accounting for the results remains an ongoing endeavour within the economic literature. One issue that is beyond debate is that human behaviour does not match the outcome predicted by the canonical economic model—at least not in the overwhelming majority of cases. In truth, this is a finding that surprises only economists, and even then it is only surprising in light of the theoretical backdrop against which the empirical results are cast (i.e. the neoclassical model of behaviour). We would stress that the game-theoretic prediction must be considered for what it is: a mathematical benchmark that specifies how players ought to behave if they make utility-maximising choices in light of the expected utility-maximising decisions of others. It is by no means an accurate model of human behaviour, not least because rationality in the economic sense is always defined in terms of pecuniary losses and gains. In light of the available evidence, one might argue that people are not strictly profit-seeking and instead rely on a more socially-oriented form of rationality. That is, economically 'irrational' decisions are actually very sensible if considered in relation to the scenario in which such decisions would ordinarily occur. For example, conforming with fairness is costly in a pecuniary sense but is highly rational in the real world, where cooperation is vital for survival in the longer term. But since people have been shown to become both more or less prosocial depending on the variable under study, evidence from the UG literature suggests that most real world decisions are guided by a mixture of self-interest and concern for the welfare of others (Camerer, 2003).

4.3.5.1 Explaining Ultimatum Game Behaviour

While it is explicitly *not* the aim of this thesis to explain why people behave as they do in ultimatum games, it now seems appropriate to consider some of the primary theoretical explanations that have been proferred within the literature. Reviewing this material helps to inform our understanding of the UG and direct our own interpretations of behaviour in the novel DLUG paradigm. Here we shall consider explanations of proposer and responder behaviour in turn.

The most straightforward explanation of UG behaviour is that offers result from fairness norms held by human players. Face value appraisal of the available evidence lends weight to this explanation, and indeed, to the average person, the idea of behaving entirely selfishly in the game may conflict with the basic standards upon which our societies operate. Nevertheless, considerable research effort has been invested in determining why people prefer fairness over the selfish option. Several explanations relate to the nature of the UG and the way in which it is interpreted by players. For instance, players may rely on equitable offers because they have a poor understanding of the game; in a one-shot setting, players are simply making naive guesses about how best to divide the source, rather than working through the problem carefully as they should (e.g. Binmore *et al*., 1985). This explanation fits well with the 'general equality algorithm' used in social decision making (Hertel *et al*., 2002), but is somewhat facile given that players are known to be able to derive the optimal solution (Güth *et al*., 1982) and exhibit preferences for fairness even after multiple rounds and learning opportunities (Roth *et al*., 1991; Knez & Camerer, 1995; Slonim & Roth, 1998; List & Cherry, 2000). In fact, one meta-analysis of UG experiments found that players actually become *more*, not less, generous over time (Oosterbeek *et al*., 2004). Bounded rationality (cf. Brenner & Vriend, 2006) presents an alternative explanation—players survey the range of available payoffs and choose an option that balances the various possible outcomes. It appears that this option is not the canonical offer, at least in the majority of cases.

An alternative explanation concerns the aforementioned threat of rejection—that proposers are afraid of rejection and, in essence, play fair to balance threat of rejection with pecuniary gain (cf. Guth & Tietz, 1988; Ochs & Roth, 1989; Kravitz & Gunto, 1992). Results from dictator games support this intuition; most show a drop in equitable allocations when compared to UG treatments. However, we saw earlier that proposers continue to make allocations above the minimum amount even when the responder has no recourse. This in itself indicates that players still see a need to behave in line with fairness norms and often find it difficult to suppress their conformity to said norms.

Various models have been proposed to account for fairness norms in economic decision making. In essence, these models aim to account for empirical findings by assuming different underlying motivations for offers and rejections. One such explanation is that of *inequity aversion* (Fehr & Schmidt, 1999). The idea here is that people generally dislike inequity and are willing to give up material payoffs in order to achieve more equitable outcomes—such an argument can account for the behaviour of both players in the UG. Fehr & Schmidt (1999) suggest that inequity aversion helps maintain cooperation; by rejecting inequity and perceived injustice, people demonstrate that they are not willing to be taken for fools, and hence lower the opportunity for others to free-ride. Alternatively, *guilt aversion* (Brenner & Vriend, 2006) represents a conceptually different argument in that proposers are assumed to experience disutility (i.e., guilt) when making an allocation that they know is selfish. In this case, fairness arises as a way of preventing this feeling; the perception of oneself as a 'bad person' is enough to sway proposers towards the equitable offer. Similarly, responders might accept a tiny offer because they feel guilty about rejection and causing 'pain' to the other. In practice, however, this latter conjecture tends not to be the case, suggesting that responders may derive some payoff from satiating anger (Andreoni & Miller, 2002), or perhaps experience a negative emotional response from perceived unfairness (Sanfey *et al*., 2003). The broader idea of moral costs in decision making also informs a model proposed by Levitt & List (2007), who emphasize that fairness may stem from a desire to balance moral costs with pecuniary gain. Other approaches emphasize the role of intentions in determining acceptance or rejection (Rabin, 1993), whereas others stress the role of altruistic concerns for allocation behaviour (Andreoni & Miller, 2002). This wide variety of explanations speaks to the inherent difficulty of accounting for the array of results obtained using the classic UG, but also emphasizes the considerable academic interest that has been invested into the model.

A final explanation that has seen a great deal of interest concerns the nature of the experiment itself; that is, as a supposed test of underlying preferences or conformity to norms, the experimental situation guides the behaviour of participants towards cooperative outcomes. Zizzo (2010) suggested that laboratory experiments are influenced by 'cognitive experimenter demand effects', including the options available to participants (List, 2007; Bardsley, 2008), unintended emphasis of variables of interest, anonymity between subjects and the experimenter (Hoffman *et al*., 1994), or the particular instruments used in the study (Larrick & Blount, 1997). For example, the mere presence of the option to give money in a dictator game might imply that donating is the correct behaviour (Winking & Mizer, 2013). Indeed, participants in dictator games actually move away from the prosocial choice when the option to steal money is introduced into the experiment (List, 2007; Bardsley, 2008). Because of these concerns, research in economics has recently steered towards capture of real-world decision making through use of natural-field experiments

(Harrison & List, 2004). Use of the 'natural' prefix here is deliberate because subjects are typically unaware that they are participating in an experiment, meaning that concerns regarding experimenter effects and anonymity are alleviated. Such experiments capture natural behaviour in tasks where 'participants' act under their own volition in the belief that they are not being scrutinised (Harrison & List, 2004).

Several such studies exist, and the methods employed are quite ingenious. Happily for our work, results suggest that behavior in the field is not that different to the laboratory. Güth *et al*. (2007) conducted an ultimatum game via the German newspaper *Die Ziet*; more than 4,500 readers 'participated' in the game by responding to a competition form which asked them to split 1,200 euros between themselves and two anonymous others. Each participant had a small chance to be rewarded with the products of their proposals. The researchers found that behaviour was not different from that of laboratory studies: 60% of respondents chose an equitable split, and this allocation was significantly more likely to be accepted by responders. As with laboratory studies, the game theoretic benchmark was the offer least likely to be accepted. Similar results were found for a newspaper dictator game study, with dictators giving about 40% of the resource to anonymous others (Ockenfels & Werner, 2012). An alternative approach was adopted by Stoop (2012) and Franzen & Pointner (2013) who conducted anonymous dictator experiments using a misdirected mail technique. Participants received an ostensibly misdirected envelope containing a cash prize for participating in an experiment. The dependent variable was whether not each individual would keep some or all of the money, or whether they would go to the effort of mailing the reward to its intended recipient. In the study by Stoop (2012), results were identical between the lab and the field: envelopes were returned about 50% of the time. In a within-subjects design, Franzen & Pointner (2013) found that dictators who had previously donated money in the lab were also more likely to return the envelope in the field.

In summary, explaining why people behave as they do in ultimatum games has proven difficult. Competing theories exist, each of which has varying levels of support and associated caveats. Whatever the underlying explanation, the key point is perhaps that the UG provides very consistent evidence regarding fair outcomes and the extent to which fair outcomes can be influenced by specific variables. Moreover, field studies indicate that findings obtained in the laboratory remain true in the real world; in some cases, laboratory studies actually underestimate the generosity shown in everyday settings. That said, counterevidence is beginning to emerge (Winking & Mizer, 2013) and thus research on explaining UG and DG behaviour is very much an ongoing endeavour.

4.4 Economic Games and Human-Computer Interaction

We began this chapter by proposing to adopt the ultimatum game to study division of labour. We now now review prior use of economic games in HCI with the aims of situating our intended contributions and considering how our DLUG paradigm could prove useful to the field.

4.4.1 Prior Use of Economic Games in HCI Research

As identified by our review earlier in this chapter, economists use experimental games in two ways. First, evidence from games has theoretical value because it presents an anomaly for the canonical economic framework, and is therefore relevant to the development of behavioural models that attempt to incorporate social preferences into utility functions (e.g. Bolton, 1991; Rabin, 1993; Fehr & Schmidt, 1999). Second, and as articulated earlier in this chapter, economic games can be used to investigate the impact of particular variables (e.g. anonymity, payoff size, gender) on the outcomes of bargaining *per se*.

Research using economic games in HCI corresponds with the latter of these two categories. This is perhaps not surprising given that refinement of economic theory is not typically a goal for HCI research. Rather, games are used to measure the impact of various situational or interface manipulations on cooperative behaviour during use of computers. Cooperation in such studies is measured by the distributive generosity of participants' allocations, with higher allocations taken as a sign of an increased willingness to cooperate. It is not always the case that these allocations are pecuniary; many studies employ tokens or virtual currencies as placeholders for the payments participants receive at the end of the experiment. The following subsections review such studies according to the subject matter they address.

4.4.1.1 Trust

The most prevalent use of economic games in HCI has been in studies of trust; in particular, how trust can be enhanced or engendered by computer interfaces. Many social cues are lost when groups have to work in distributed and computer-mediated settings (Olson & Olson, 2000), and loss of such cues can make it difficult for team members to establish mutual trust, especially if they lack information about their collaborators' actions, intentions, and emotions (Riegelsberger *et al.*, 2003).

Work in this space argues that the tendency for individuals to cooperate or defect is a proxy for the level of trust among group members; groups with higher collective gains are understood to have established greater trust, with high defection rates viewed as a sign of distrust (cf. Bos *et al.*, 2002; Scissors *et al.*, 2009). An early study by Rocco (1998) used a public goods game to explore trust formation in face-to-face and electronic settings, finding that people were more trusting in the former setting. Other studies have employed the prisoners' dilemma to assess the impact of communication medium (Jensen *et al.*, 2000; Zheng *et al.*, 2001) and personal online profiles (Davis *et al.*, 2002) on trust development. More recently, the 'Daytrader' social dilemma task, a variant of the prisoners' dilemma where participants choose how to invest earnings over a series of rounds, has been employed to study the impacts of communication medium (Bos *et al.*, 2002; Zheng *et al.*, 2002; Nguyen & Canny, 2007) and linguistic similarity (Scissors *et al.*, 2008, 2009) on computer-mediated trust at the intra-group level.

4.4.1.2 Social Distance

Studies of distributed work have shown that 'distance matters' (Olson & Olson, 2000) and that the social distance between coworkers can be a barrier to collaboration. Bradner & Mark (2002) studied social distance using an iterated prisoner's dilemma with six rounds. They found that participants located in Irvine, California, initially preferred to cooperate with a partner situated nearby over a partner on the eastern seaboard (in this case, Boston, Massachusetts). More design-oriented work has sought to address the problem of social distance through awareness mechanisms. Kim *et al*. (2012) report a study where groups participating in a public goods problem were provided with sociometric feedback about the behaviour of their collaborators. The sociometric feedback successfully increased the level of cooperation in groups who began the task in distributed settings.

4.4.1.3 Agents and Avatars

Finally, economic games have been used in studies exploring cooperation between humans and computerised agents. Early studies by Kiesler *et al*. (1996) and Parise *et al*. (1996) employed the prisoner's dilemma to study willingness to cooperate with agents represented by human faces. Both studies found that people were more willing to cooperate with increasingly human-like computers, implying that interface agents might be more successful if they possess human-like qualities. More recent work has developed this theory by designing algorithms that can account for human fairness preferences during interaction with agent-based systems (de Jong *et al*., 2008).

Research has also used economic games to explore how people behave in virtual environments. Yee & Bailenson (2007) used the ultimatum game to study the impact of varying the appearance of a user's avatar.[15] Participants played four rounds of the UG with a confederate who, in this study, acted as the responder. In the first condition, participants used an avatar that was noticeably taller than the confederate's. In the second cell, the participant's avatar was noticeably shorter. The researchers found that those in the taller avatars were more likely to make and reject unfair offers, the implication being that people act in line with their stereotypical perceptions of tall people while under the guise of a taller avatar. Finally, Lazem & Gračanin (2010) and Lazem *et al*. (2012) used a commons dilemma to examine communication in the virtual world of *Second Life*. They found that groups who were allowed to communicate via voice were better able to sustain the common resource pool, as compared with groups who relied solely on the visual actions offered by their avatars.

4.4.2 Merits of our Approach for HCI

Having considered prior use of economic games in HCI and CSCW, we propose that our adaptation of the UG has merit for the following reasons.

A majority of prior studies in HCI have simply applied economic games in extant form—no work has used economic games to study division of labour in the manner we have proposed. Some studies conducted outside of HCI have explored negotiation over other incentives, e.g., waiting time in a UG (Berger *et al*., 2011), yet no research has attempted to explore a variable quantity of work as the payoff in a bargaining situation. The use of pecuniary incentives is assumed to induce known preferences in subjects, i.e. that

[15] An avatar is a human-like digital representation of a user in a virtual environment.

more money is preferred over less (Smith, 1982). However, many social transactions in everyday life are one step removed from financial payoffs, and this may cause behaviour to change (Ariely, 2008; Berger *et al.*, 2011). Within economics, the use of field goods is seen as a progressive exercise over induced valuations (Harrison & List, 2004; Winking & Mizer, 2013). Such studies speak to the potential impact of commodities on fairness; for example, Takahashi (2007) found that persistent smokers were tolerant of inequality when bargaining over money but not over cigarettes. In line with such work, our DLUG allows us to study the division of a currency held as common interest by those in the field of CSCW—that is, *work*, and the coordination of work undertaken by people in pursuit of collaboration and cooperative activity. This has utility for CSCW because it allows us to address an important question for collaboration: to what extent do people reason about workloads in the same way that they reason about other resources? Such a contribution is made possible by both the design of economic games (their simplicity and the fact that they can be so readily transfigured) and the substantial literature reporting reliable phenomena concerning the way they are played.

We recognised earlier that one strength of economic models is their simplicity. However, this cleanliness comes at the expense of scope; the UG and other economic tools are unavoidably narrow in terms of how they model cooperation. In the UG, players come to the lab, participate in one round of negotiation with an anonymous individual, and then leave with their payoffs. There is thus no anticipation of future interaction, a situation that may be constructed for good scientific reasons (e.g. to minimise the impact of reputation) but is arguably not representative of the real world. Our DLUG compensates for this by isolating the planning phase of work, with players then required to complete the procedural outcome of actually *doing* their allocated tasks. We feel this offers several opportunities. First, it pushes the decision scenario closer to the real world endeavour of doing collaborative work—any such decision about division of labour around a group is made in the knowledge that one has to work with the individuals in question, even if this work is neither co-located nor concurrent. Second, as an exploratory scientific endeavour, the DLUG provides an opportunity to study fairness preferences in both allocation and enactment. This latter conjecture is warranted by an emergent 'matching' effect, as reported in the following chapter.

Finally, as we design and build systems that support collaboration on a large scale, understanding aggregate effects associated with fairness and justice in work will be increasingly important (Benkler, 2012). In particular, the volume of anonymous, computer-mediated collaborative work is likely to increase with future developments in crowdsourcing, microwork, and other mass-contribution systems (Kittur, 2010). Models such as ours, that allow for the exploration of cooperation and preferences in the allocation and enactment of work, could be used to inform division of labour policies in such systems. There is also no reason why the aforereviewed studies using economic games in HCI could not be replicated using the DLUG—such studies might unveil new insights about how variables influence allocation of work over tokens or pecuniary payoffs.

4.5 Chapter Summary and Conclusion

In this chapter we first proposed to adapt the classic ultimatum game to study division of labour. We presented our novel division of labour ultimatum game and proposed to explore it in an empirical setting. To provide a proper theoretical and methodological foundation for our own model, we reviewed relevant empirical literature on the the ultimatum game and identified a range of factors that must be controlled in our own experiments. Our review showed that what is regarded as a fair outcome in the UG can be shifted according to myriad factors. This further emphasized the subjective and variable nature of fairness but provided a clear demonstration of how fairness can be explored using economic games. We then explored how the ultimatum and other games have been used in prior HCI research in order to articulate the potential impact of our own contributions. The following chapter reports the findings from a series of empirical studies where our DLUG was iteratively explored in an experimental context.

CHAPTER 5

EMPIRICAL STUDIES USING THE DIVISION OF LABOUR ULTIMATUM GAME

5.1 Chapter Overview

Chapter 4 introduced our novel division of labour ultimatum game (DLUG) and presented a select review of the theoretical and empirical literature associated with its predecessor, the classic ultimatum game (UG). In this chapter, we present a series of studies where the DLUG is employed to investigate fairness in the division and completion of collaborative work. To this end, we instantiate our game with the task of collaborative information seeking, a nascent area of research in which division of labour is a relevant research question (Foley & Smeaton, 2010; Kelly & Payne, 2013). Our approach is incremental; an immediate aim is to explore behaviour in the DLUG, and our results provide further evidence of fairness preferences in participants' distributive allocations of workload. We also study participants' work process, and through analysis of behavioural protocols we identify strategies used to coordinate work while avoiding redundancy. We suggest that these strategies could be employed to enable divisions of labour in specialised collaborative search tools. Additionally, we report an emergent finding concerning overall equality in participants' search times. This finding is highly unexpected, and we attempt to interpret and explain the result by drawing on relevant work from social psychology.

This chapter is organised as follows. We briefly reiterate our DLUG as a model of articulation work and provide justifications for our decision to focus on web-based collaborative information seeking. This allows us to pinpoint the way our studies could usefully contribute to the literature on collaborative search. We then describe how the present studies have been designed to account for the factors known to affect ultimatum games, as outlined in Chapter 4. We then present our experiments, which focus on three types of protocols: allocations and participants' stated rationale for said allocations; coordination strategies; and an analysis of participants' task completion times. The chapter ends with a general discussion that attempts to consolidate each finding alongside relevant literature.

5.2 Foundations for the Present Studies

To reiterate, we propose to explore division of labour using our novel twist on the ultimatum game. In our game, two players are tasked with agreeing a division of a known quantity of work items. Completing the work provides each individual with a fixed pecuniary reward. If the players can agree an allocation, the two are allowed to collaborate and contribute work items to a common pool. However, if they cannot agree, they must each work individually. The game begins with one player, the proposer, suggesting a division; the responder then accepts or rejects the proposal. In the event of acceptance, players collaborate and complete the task according to their agreed allocations. If the second player rejects, however, then the game ends and players must work alone.

The subsections below outline some important decisions regarding our task choice for exploring the DLUG. We also explain how the various factors identified in Chapter 4 have been accounted for and controlled in our experiments.

5.2.1 Task Choice: Web-Based Collaborative Search

Our DLUG necessarily requires a collaborative task in order to become operational. One of the aims of the present work is to demonstrate the theoretical and practical merit of the DLUG for HCI. Therefore, we chose a task that allows us to investigate the allocation of work while providing sufficient opportunity to acquire implications relevant to interface design. Our task of choice is web-based collaborative information seeking, also known as collaborative search. We use these terms synonymously because they refer to the same behaviour: that of two or more "participants work[ing] together to satisfy an information need" (Morris, 2013, p. 1182). Collaborative search is a nascent research topic that has seen growing interest within the HCI, CSCW, and information retrieval communities. To narrow our efforts, we characterise collaborative search as involving at least two searchers conducting *explicit, intentional* work (cf. Golovchinsky *et al*., 2011). This definition precludes consideration of recommendation and filtering systems that algorithmically enhance search using prior results from anonymous 'collaborators' (e.g. Park & Pennock, 2007).

As a test case for our model, collaborative search is a good fit because it is a realistic situation where collaborators must decide how to partition the task while defining a management strategy to prevent redundancy in their information seeking activities (Foley & Smeaton, 2010). In other words, collaborative search involves both quantitative allocation and coordination of work, thereby capturing both aspects of the phenomenon we wish to investigate. In addition, division of labour is a relevant research problem in the literature on collaborative search (Foley & Smeaton, 2010; Golovchinsky *et al*., 2008a; Morris, 2008; Kelly & Payne, 2013). Several authors have designed and developed specialised systems to support collaborative search. Examples of such tools include *SearchTogether* (Morris & Horvitz, 2007), *WeSearch*

(Morris *et al*., 2010), and *CoSense* (Paul & Morris, 2009). (Further systems are described in Chapter 6 of this thesis.) Division of labour is one concept underlying the design of these systems (Morris & Horvitz, 2007; Foley & Smeaton, 2010) but little is known about how people organise their efforts during collaborative search without specialised system support. This gives us the opportunity to contribute to current literature by investigating how searchers coordinate when carrying out their experimental task. Identification of any coordination strategies would provide immediate implications for design—future collaborative search systems could be designed to partition work in accordance with said strategies.

5.2.1.1 Division of Labour in Collaborative Search: Systems and Approaches

Within the literature on collaborative search, four general approaches to managing division of labour have been identified (Golovchinsky & Pickens, 2009; Kelly & Payne, 2013), each of which involves a different degree of specialised support from the system. These are:

Communicative approaches, where people exchange information about their search activity in order to avoid redundancy and coordinate work. This approach entails no specific system support for division of labour beyond methods of exchanging information, e.g. chat functionalities (González-Ibáñez & Shah, 2011)

User Interface approaches, where tailored support for division of labour is provided at the user interface. *SearchTogether* (Morris & Horvitz, 2007), for example, includes buttons that allow collaborators to distribute results from individual queries or specific search engines around the group.

Algorithmic approaches, where data about search activity is collected at the system layer and is re-used to enhance information seeking. For example, algorithms can be designed to surreptitiously remove results from a user's queue according to prior judgements of relevance from collaborators (Foley & Smeaton, 2009).

Role-based approaches, which prescribe the assignment of tasks in accordance with roles that have predefined duties. For instance, the *Cerchiamo* search system requires one user to spend time finding sources while another evaluates findings and provides feedback on their relevance (Golovchinsky *et al*., 2008a).

Research suggests that each of these approaches can be beneficial for collaborative search, though each does have its associated caveats (see Kelly & Payne, 2013, for a review). However, relatively little is known about how collaborative searchers coordinate in the absence of technological support. If emergent practices were identified, future systems could be designed to support such behaviours. Prior work by Morris (2008) identified two strategies used by collaborative searchers to divide work in the absence of specialised support: *divide-and-conquer*, where searchers explicitly coordinated their activities by using different search engines or keywords, and *brute force*, where searchers did not coordinate their activity and instead coped with redundancy when merging their results later on. However, the existence of these two strategies arises from searchers' self reports rather than direct empirical observation. The present studies extend this literature by identifying other strategies used by searchers when coordinating collaborative search. We argue that these strategies could be supported within the design of collaborative search tools.

To further scope our concerns, the present experiments are constructed to consider only distributed, synchronous collaborative search. We realise that focusing on one instance of collaborative work constrains the applicability of our findings from the present studies, at least to a certain extent. However, focusing on one instance of collaboration allows us to hold parameters including interdependency, i.e., tight versus loosely coupled work, and temporality, i.e. synchronous versus asynchronous work, as constant. We fully expect that the frequency and complexity of interactions will influence the division of labour and resultant needs for coordination (cf. Olson & Olson, 2000) yet we hold these constant to facilitate inter-experiment comparisons. Our results are then available for comparison with future studies using different tasks and experimental setups.

Finally, it seems important to consider whether our task is sufficiently rich to be characterised as 'collaboration'. As will be seen in our experimental procedure, participants' task is to collect sources from the Internet to form a shared bibliography. Such a design is intended to mirror the early phases of shared literature review, a common collaborative information seeking task (Morris, 2008, 2013). In our experiments, participants complete this task through a chat client, searching for and contributing sources on-the-fly until the required total is reached. This act of 'shortlisting' also matches the process adopted during real-world collaborative information seeking (Kelly & Payne, 2014, see Chapter 6 of this thesis). While our experimental conceptualisation of collaborative search is necessarily somewhat reductionist, we believe that it involves behaviours necessary to satisfy a basic definition of collaboration (such as that outlined in Chapter 2 of this thesis). Our task involves division of labour, information sharing, and contributions towards a common goal, all of which are integral aspects of collaborative work. We see our approach as an almost 'minimal collaboration paradigm', in that we capture the basic processes involved in collaboration while abstracting away from potentially influential factors, e.g. physical appearance, spatial setup, and group dynamics more generally (Levine & Moreland, 1998).

5.2.2 Remarks on Experimental Methodology

5.2.2.1 Controlling Factors that Impact Behaviour in Games

As evidenced by our review in Chapter 4 of this thesis, ultimatum games are sensitive to many aspects of experimental design. This subsection remarks in detail upon how we have considered such issues in the design of our experiments. These considerations are important because, as this is the first study of its kind, it is important that we establish a consistent protocol for our experiments while mitigating the impact of influential variables. This will allow for legitimate comparisons. The decisions reported below are true of all experiments reported in this chapter, unless otherwise stated. The most pertinent issues and our methodological choices were as follows:

Anonymity: we designed the present studies as simple collaborative situations where participants had total anonymity from one another. Not only does enforcing anonymity follow the procedure typically used in behavioural decision making experiments (Camerer, 2003; Croson, 2005), keeping anonymity constant controls for the impact of identifiability itself; participants are known to become more or less generous when anonymity is relaxed or tightened (Hoffman *et al.*, 1994; Bolton & Zwick, 1995; Bohnet & Frey, 1999; Haley & Fessler, 2005; Wilson *et al.*, 2010). This also has the dual benefit of controlling for gender effects and physical attractiveness (Schweitzer & Solnick, 1999; Eckel & Grossman, 2001; Solnick, 2001; Engel, 2011). Although we recognise that, in the real world, most collaborators are likely to know each other and search tasks will not always be completed under conditions of anonymity, we would argue that to understand the effects of interpersonal knowledge requires some understanding of how collaboration is managed in its absence.

Repetition: we opted to use a single round of the DLUG, with participants then going on to complete the task as agreed. While it is true that repeated play is more realistic, in the sense that real-world negotiations are more likely to be iterative rather than one-shot, proper understanding of such play requires some knowledge about behaviour in one-shot games (Forsythe *et al.*, 1994). Our paradigm is novel and thus it seems appropriate to begin at the most basic level before proceeding forwards. One-shot games also control for punishment strategies and reputation effects (Roth *et al.*, 1991; List & Cherry, 2000), both of which could conceivably impact division of labour.

Communication: participants in our experiments were not allowed to communicate prior to completing the DLUG. This minimises the impact of 'cheap talk' (Croson *et al.*, 2003) and the positive effects of communication more generally (Sally, 1995; Frey & Bohnet, 1997; Bohnet & Frey, 1999) on allocations. The only communication that occurs during the allocation is the offer and the response, both of which are made via computer terminals (more detail is presented in the description of experiment one). We then relax the constraint on communication during the process of completing work, such that participants are allowed to communicate, coordinate, and share findings during their collaborative search task. Such a design holds greater external validity in the sense that arbitrary constraints on communication during the enactment of work would be unrealistic.

Stake size: we saw earlier that stake size does not have much of an impact on ultimatum play, in that people continue to make equitable offers even when large amounts of money are on the table (Forsythe *et al.*, 1994; Hoffman *et al.*, 1996a; Slonim & Roth, 1998; Cameron, 1999; Oosterbeek *et al.*, 2004). While we can intuitively expect that our participants will experience some disutility if asked to complete more work for less pay, we cannot give an exact *a priori* estimation of how varying the quantity of work will impact allocations. As such, we opted to use a payoff of 10 work items as a straightforward mapping of the typical $10 stake used in ultimatum games. This was not an arbitrary decision: we considered that the task of finding 10 sources was a reasonable request of our participants given expected experimental constraints (trials < 1 hour) and anticipated financial payment (£5).[16] This work inevitably raises questions about the subject of reduced workload as an incentive, and we return to this matter when discussing the results of these experiments.

Entitlement: we randomly assigned participants to the roles of proposer and responder to avoid creating perceived role or resource entitlements (Güth & Tietz, 1986; Hoffman *et al*., 1994; Cherry, 2001; Cherry *et al*., 2002; Oxoby & Spraggon, 2006). Our participants must earn their monetary payoff by completing some work, but there is no initial reason for either player to do less work towards this goal, i.e. we do not 'suggest' particular allocations. This ensures cleanliness in terms of testing decisions about division of labour.

Framing: we controlled for framing effects by keeping instructions consistent for all of the present experiments. These were adapted from instructions used in prior UG experiments (Forsythe *et al*., 1994; Hoffman *et al*., 1994). We were careful not to imply any need to share or be fair, and did not make any major changes between experiments that could explain shifts in behaviour, e.g. by changing the language from 'dividing' to 'allocating' work (cf. Larrick & Blount, 1997). All workload units were described in absolute numeric terms, rather than percentages (Croson, 1996).

Demographics: we saw earlier that gender, academic background, and culture can all impact ultimatum bargaining. The impact of gender is mitigated by our decision to keep participants anonymous. While the impacts of education and culture could be controlled in theory, this would be restrictive as we do not have easy access to large numbers of participants from a particular creed or culture. Instead, we had to rely on self-selecting volunteers. We therefore relaxed this constraint and allowed people from across our campus to participate. Since our participants had a diverse array of academic backgrounds, any effects associated with a particular discipline should be minimised.

To an extent, some of these decisions were tempered by our desire to balance the established methods of experimental economics with a reasonable construction of collaborative work. Consider the role of communication; in economics, it is highly atypical for participants to communicate, whereas communication is an integral aspect of collaboration. Disallowing communication would, therefore, be welcomed by the former but unjustifiable for the latter. This is exemplary of the complexities raised by the present work. In any case, our efforts to hold these key parameters constant, and avoiding arbitrary changes from one study to the next, should allow the results of the present studies to be compared to any future work that takes place in this area. Our established method can be used in future replications and studies that examine particular variables, e.g. by increasing the total workload or by relaxing anonymity between participants.[17]

[16]As it transpired, sessions took between 15–50 minutes to complete, but participants were not told how long sessions would last prior to recruitment and thus had no expectations about the amount of work involved. It is possible that such explicit knowledge about how long a particular task will take may affect allocation behaviour. We will return to the issue of incentives during the general discussion of these studies.

[17]Future work could also draw on these variables to bring the study closer to the real world, e.g. by allowing communication before and during the allocation process. (*"What would you do if I made this offer?"*, or *"I don't like that offer, make it lower and I'll accept"*).

Having explicated these issues, we now move on to exploring behaviour in the DLUG. Our first experiment was intended as a pure instantiation of the game, where collaborators have limited knowledge about their task and are bargaining over notional work items.

5.3 Thesis Study 2: Division of Labour as an Economic Game

5.3.1 Design

We designed our first DLUG experiment as a simple situation in which participants were required to complete an information seeking task to earn a fixed payment of £5. The exact task was to use a Web browser to locate 10 sources so as to form a bibliography that addressed the following question: *"To what extent can design be considered a psychological process?"*. This question was selected on the basis of pilot trials which indicated that the question was understandable to those with little knowledge of the topic but still sufficiently broad that the task was not trivial. We opted for a bibliography task because collection of papers and webpages is a common and realistic collaborative search task (Morris, 2008). We did not provide participants with any specialised support for this task beyond a web browser for search and a chat program for communication. As hinted earlier, this allows exploration of the work management strategies used by participants in the absence of any particular technological encouragement or facilitation.

Rather than simply asking for a list of sources, we presented the task description as 'forming a reading list of 10 reliable sources for students of a night class'. This was done so as to introduce a soft yet meaningful criterion to the search task, which we hoped would encourage participants to evaluate sources more carefully before pasting them into the chat window. Participants in the aforementioned pilot trials (which did not initially include the description emphasising reliability) often chose to paste the first 10 references that appeared in Google. This behaviour is undesirable in the sense that it requires little effort to coordinate the work, in turn making our observations of the work process somewhat meagre. We found that introducing the reliability criterion encouraged participants to invest more effort in evaluating the quality of sources prior to inclusion.

While our task description suggested that sources should be 'reliable', we opted not to specify exact criteria for what constituted a suitable source for the reading list. We also withdrew from specifying exact requirements for effort and time on task. These are all factors which could affect the allocation behaviour, and thus rather than constrain them we allowed participants to self-regulate their effort investments and judgements of relevance. We hoped that this would allow for interesting and naturalistic behaviours to emerge, and indeed this decision was warranted by our process results where we unveiled novel coordination strategies and an emergent effect in task completion times.

For this study, participants were not told about the search topic before dividing up the work. Although this has less external validity in terms of real-world work allocation, this design represents a purer instantion of the DLUG, in terms of bartering over work as a currency, and allows investigation of allocation behaviour in advance of specific task knowledge. (The issue of task-related knowledge is addressed head-on by our next experiment). Pairs of remote participants thus agreed a division of ‘work items’ through the DLUG, and then completed the task according to their agreed division. Participants coordinated their activity through an instant messaging (IM) program and all activity in each client was logged and timestamped for later analysis.

We also considered the opportunity cost associated with participating in the present experiment. Essentially, the work requested of participants was a barrier to the collection of their £5, which was presumably their sole motivation for volunteering into the experiment. It is possible that seasonal changes could affect participants’ desire to do the work required and, in turn, lead them to want to leave more or less quickly (cf. Berger *et al.*, 2011). In summer, for example, participants may be especially keen to enjoy the hot weather, meaning they could make unfair offers in order to escape the experiment sooner. On a rainy day, participants might be perfectly content to remain in the lab. While this is an inherently difficult issue to control for, the study reported here was conducted over a three-week period in August 2010, meaning that seasonal effects should not be a major impact. Trials were conducted at one of three intervals (10am, 12pm, and 2pm) and no more than two trials were conducted on any particular day.

A final consideration was whether or not the task of pasting references into a shared bibliography was sufficiently collaborative to produce meaningful results, i.e. whether or not it really constitutes ‘collaboration’. Given that, when completed collaboratively, our tasks have participants who are explicitly aware of their collaboration, have shared intentionality, offer contributions to the common good, and have some vested interest in the outcome of the work, we believe that our tasks can satisfy our earlier definition of collaboration presented in Chapter 2.

5.3.1.1 Hypotheses

Since we did not design the present experiment with specific hypotheses in mind, and were not aiming to test any particular variable, we eschew exact predictions concerning how participants may or may not behave. Moreover, this is the first experiment using the DLUG, and thus behaviour is currently an open question.

Recall, however, that there are at least two naive predictions about how players might behave. First, there is the baseline prediction of self-interested behaviour, which would see proposers allocate the maximum possible amount (9 items) of work to their anonymous partner. The responder should then accept this offer as it confers a 10% reduction in workload. However, there is a second intuition; based on the fact that players desire equity and anticipate rejection, offers in ultimatum games typically converge at the point of an equitable split. It is possible that the same actions might emerge in the DLUG. These issues are worth keeping in mind when appraising the results reported here.

5.3.2 Materials

Participants used a Microsoft Instant Messenger (IM) client to play out the DLUG and subsequently communicate during the Web search task. All searches were conducted using the Mozilla Firefox web browser. A non-intrusive Firefox extension, *HCI Browser* (Capra, 2010), was used to record search behaviour surreptitiously, including queries and page visits, into a text file.

To obtain informed consent, we used a paper-based script designed in accordance with our institution's ethical guidelines. Copies of all instructional materials used in this experiment can be found in Appendix C. We used two sets of instructions to present our experimental task. The first page was given to participants prior to the DLUG, and presented the task as follows: *In this study you will be required to search the Web for information in order form a reading list for students of a night class. The reading must have 10 reliable sources.* The instructions then explained how to play out the DLUG via the messenger client. The second page of instructions described the web search task, presented as follows: *You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information that addresses the following question: "To what extent can design be considered a psychological process?*.

We used a post-experiment questionnaire to gauge computer expertise and search experience using measures from Morris (2008). To probe decision making about allocation behaviour, we included the following questions: 1. *Why did you choose to allocate the work in the way that you did?* 2. *What factors, if any, influenced your decision?* 3. *Did you consider making a different offer? If so, what did you consider?*. We also included a 5-point Likert scale to gauge participants' familiarity with the search topic (1 = no familiarity, 5 = high).

5.3.3 Participants

Nine female and 11 male students and research staff from six different University departments took part in the study. Participants' ages ranged from 20–44 years (M = 28.3 years). Assignment of participants to pairs was a matter of timetabling convenience. Three pairs were male-male, two were female-female and all others were mixed pairings.[18] As mentioned earlier, we offered participants £5 as an incentive to participate in the study.

All participants reported conducting Web searches daily. Thirteen participants reported 'expert' computing skill, with the remaining seven reporting 'intermediate'. These findings indicated that our subjects were comfortable with Web search and none were novice computer users, who we might expect to find exploratory search quite difficult. The topic of design psychology was not familiar to most of our participants (M = 1.8, SD = 0.9).

5.3.4 Procedure

Participants completed the study in pairs, and were located in separate offices in different areas of our campus. These offices were separated by about 200 feet, thereby minimising the chances that participants would meet. At no point were the participants introduced face-to-face before, during, or after the study. Care was taken to maintain anonymity in each trial by using gender-neutral language when referring to the

[18] We provide this information to show that gender was distributed across conditions and roles, i.e., not all proposers were male.

other participant, e.g. "we are waiting for *the other person* to arrive". Sessions took from 15 to 50 minutes to complete and no time limits were imposed on participants.

On arrival at the study, participants were assigned to the role of either proposer or responder. Each participant was introduced to the study by a facilitator. Facilitators were connected via the Facebook chat client, allowing correspondence about participant arrival, study completion, etc. Participants were told that the study was exploring the organisation of workload during a Web search task. After providing informed consent, participants were told that the study had two parts and that each part would be explained in turn. Participants were first handed the division of workload instructions. Instructions were worded carefully, and modelled on those used in previous ultimatum game studies (Forsythe *et al*., 1994; Hoffman *et al*., 1994; Larrick & Blount, 1997) so as to avoid any experimental demand implying the need to share or be fair.[19] Each participant was told that they had to form a bibliography containing 10 sources in order to complete the study and receive their payment. However, they had the opportunity to collaborate and divide their workload, such that any sources obtained would contribute jointly to a shared bibliography of 10 items. In line with our experimental design, participants were not told the exact search topic but only that they would be forming a bibliography. Proposers were asked to specify an allocation by typing *'I want you to find χ sources'* into the chat, where χ was to be replaced with their numerical allocation. Responders then had the option of accepting or rejecting the offer in the same chat window by typing either 'accept' or 'reject'. In the event of acceptance, participants would form a joint bibliography and amalgamate sources as they were found by pasting them into the chat window. However, in the case of rejection, both participants would have to find 10 distinct sources separately.

After allocating work via one round of the DLUG, participants were introduced to the Web search task. We asked participants to record their found sources by copy-pasting a snippet of relevant text into the chat, alongside a hyperlink to the source. Participants were told that they could use any search engine to find sources, which could be anything they deemed relevant to the topic (webpage, blog, scholarly journal, etc). Beyond the requirement of 10 sources, the only other criterion given to participants was that they could not post sources their partner had already found, i.e., all sources had to be distinct. Participants could communicate via the chat during the study, but were asked not to reveal any personal or identifying information. Participants were asked to notify the facilitator once they had completed the task, at which point they were presented with the post-study questionnaire. They were then free to leave the experiment, irrespective of whether or not their partner had finished.

5.3.5 Results

Our results focus on four metrics. We present the outcomes of the DLUG first, alongside participants' reported motivations for making allocations. We then examine participants' strategies for coordinating work and preventing redundancy. Finally, we examine participants' paired task completion times so as to gauge their level of investment in our experimental task.

[19] Copies of the instructions used in this experiment can be found in Appendix C.

Session ID	Proposer's Allocation	Responder's Workload	Outcome
1.1	5	5	A
1.2	4	6	A
1.3	5	5	A
1.4	5	5	A
1.5	5	5	A
1.6	5	5	A
1.7	6	4	A
1.8	5	5	A
1.9	5	5	A
1.10	5	5	A

Table 5.1: Workload allocations and acceptances, study 1, *A = Accepted, R = Rejected*.

5.3.5.1 Workload Allocations

Table 5.1 shows the allocations made in each trial of this experiment. All of the workload proposals made were accepted, i.e. there were no rejections. Eight proposers offered an even (5-5) workload split. Two proposers made uneven splits; the first delegated six sources to her partner and four to herself (4-6), whereas the second did the opposite, allocating just four sources to her partner and six to herself (6-4). Interestingly, both 6-4 allocations were implicitly renegotiated to the point of fairness during the process of work. In other words, despite the agreement, each participant posted exactly five sources. In both cases, the partner in charge of four sources completed their share of the work first and spontaneously went on to obtain and post a fifth source. This occurred in spite of the fact that participants knew that they could leave after finishing their own portion of the task.

5.3.5.2 Motivation for Dividing Work

The post-experiment questionnaires revealed that six proposers who made a 5-5 split invoked fairness as a complete explanation of their offer. Two of these proposers revealed that they had wanted to make an unfair split but were reluctant to do so because of the threat of rejection. One cited efficiency in the negotiation process: *"simplest possible split, no need for negotiation, no risk of rejection"*. Another saw a fair division as the most efficient way of completing the task as quickly as possible due to the synchronous work situation. The explanations for the uneven offers were not, in the author's opinion, totally convincing. The generous 6-4 proposer cited the same fear of rejection as two of the fair offer proposers, whereas the ungenerous proposer claimed: *"I enjoy delegation"*.

5.3.5.3 Coordination Strategies

The major coordination problem faced by our participants was avoiding redundancy, i.e. duplicated sources. Since we did not instruct participants to overcome this problem, we examined the IM chat logs in an attempt to explore the naturalistic strategies employed to ensure coordination and prevent redundancy. We identified two clear strategies: Wait & Repair, and Partitioning by Web/Document Space.

Wait & Repair. *A priori*, one strategy for dealing with duplications would be to simply wait for them to arise and then repair them, either immediately (in the way a repair works in conversation (Clark & Schaefer, 1989)) or at a later combination stage (as in the *brute force* strategy identified by Morris (2008)). Such a strategy, which we deem *Wait & Repair*, will only be identifiable in situations where duplications actually do occur. This happened in only one pair, and the repair was managed immediately, as shown in the following transcript from the chat logs:

> [Pair 1.2]
> Participant Two: I think we found the same *[pertaining to Participant One's last source]*
> Participant One: oki
> Participant One: I will look for another
> *[Participant One pastes a reference and snippet of text.]*
> Participant Two: Good job!

A further seven pairs made no explicit mention of strategy and never suffered a duplication. Since it is likely that these pairs would have spontaneously repaired breakdowns if they had occurred, we have provisionally classified these as examples of the Wait & Repair strategy.

Partition Web/Document Space. A second available strategy for avoiding duplications is to explicitly coordinate work by searching for different types of source, perhaps by using different Web services. We observed two pairs discussing and following this strategy. In one case, the discussion emphasised document types (scientific articles versus blogs and websites), and in the other, it emphasised web services (Google scholar versus EBSCO). However, there is a natural confound between services and document types, and indeed both discussions reflected this, so for now we prefer to consider these as versions of the same strategy. We title this strategy *Partition Document/Web Space*. The vignettes below illustrate pairs dividing the task in this way:

> [Pair 1.8]
> Participant Two: I'll use Google scholar to find scientific papers. Is that ok?
> Participant One: sounds great, and I will try to find information from news & blogs
> Participant Two: perfect

> [Pair 1.10]
> Participant One: How about you look at non-journal sources and perhaps news and I take google scholar, ISI and EBCSO?

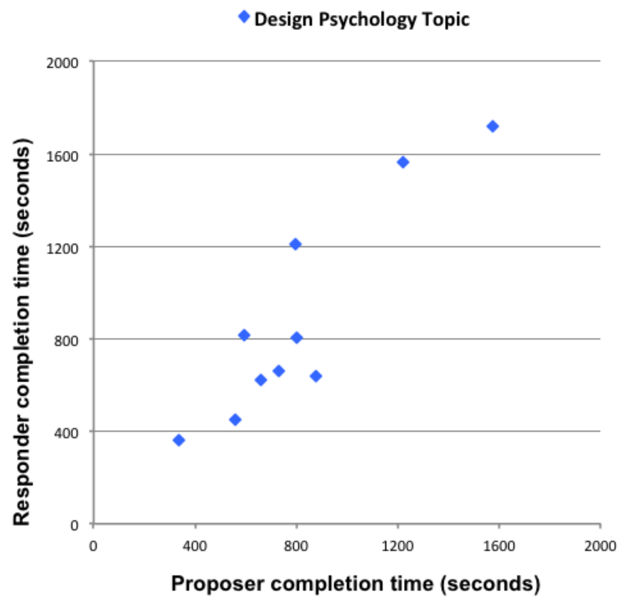

Figure 5.1: Correlation between task completion times of proposers and responders, study 2, r_s = .77).

> Participant Two: why don't we both try to find what we find first and maybe share later? not sure there's much around outside non-journals.
> Participant One: Ok I am going to work in ISI then..searching for 'psychological processes design'

5.3.5.4 Task Completion Times

We obtained an overall search time for each participant, i.e. the elapsed time between the point at which the first query was issued by one member of the pair and the time each searcher pasted their final source into the chat. Our motivation for obtaining this data was twofold. First, we wanted to gauge participants' general level of investment in our experimental task. Second, informal reports from facilitators suggested that pairs were finishing at similar times, but that there might also be some between-pair variance. Analysis of the search time data allowed us to explore whether this observation carried any weight.

Eyeballing the data, we observed that within-pair completion times were strikingly similar, but that there was also a clear difference *between* pairs. This dual effect is illustrated in Figure 5.1, which graphs the raw search time data for the experiment. In order to examine the strength of the trend and assess its statistical significance, we first transformed the data using natural logarithm in order to remove positive skew (Skewness statistic = 0.942). Natural logarithm was used because logarithmic transformations are preferable when data is positively skewed (Sheskin, 2011), are appropriate for ratio level data like time (McDonald, 2008), and are necessary in order to normalise the data to better meet the assumptions of statistical tests (McDonald, 2008; Sheskin, 2011). A correlation was then computed between the paired search times using Pearson's product-moment correlation coefficient. We observed a strong positive correlation, $r(8) = 0.88$, $p < 0.001$.

5.3.6 Discussion

This experiment sought to explore our empirical model based on the ultimatum game. We asked participants to bargain over the allocation of work items in a collaborative information seeking task. The results are indicative of fairness preferences in the division and completion of labour.

At a broad level, behaviour in the DLUG mirrored findings from the standard UG in that participants did not appear to follow a behavioural model based on pure self-interest. Studies of the UG have consistently found that the modal offer tends to be a fair split of money (see, e.g. Camerer & Thaler, 1995; Camerer, 2003), and the same is true of the experiment reported here. We found that eight of our 10 offers were 5-5, with the remaining two close at 4-6 and 6-4. That participants did not attempt to free ride by minimizing their own workload is indicative of a tendency towards fairness in the division of labour. Indeed, although our sample size was very small, the prevalence of completely fair allocations is even stronger than typically observed in the economic literature, where the modal outcome of fairness tends to account for about 50-60% of a given sample.

It is also interesting that workloads were managed so as to be even more explicitly equal during the process of completing work. This equality was realised in two ways. First, the rare cases of unequal allocation were implicitly reworked during the course of completing the task. This suggests a preference for correcting initial inequity during the process of work, perhaps due to an implicit form of positive reciprocity, whereby the party with the smaller workload takes on some of their partner's work in order to thank them for generosity shown during the workload division.

Second, and perhaps more interesting, was the finding of a strong, positive correlation between paired search times. One part of this result relates to the high degree of similarity *within* a pair. This is interesting in light of the fact that there appears to be little reason for participants to finish at almost the same time. If anything, the circumstances of the experiment, where participants have minimal awareness, work anonymously, and can leave after completing their individual allocations, might be expected to count against this behaviour. What makes the finding doubly interesting is that this similarity is consonant with a clear difference *between* pairs, even though all pairs work on the same task. This seems to indicate that pairs somehow synchronised, albeit implicitly, their overall work rate so as to finish at similar times. At this juncture we shall not elaborate on this effect since we have a low number of participants—we feel that more data is required to explore whether this effect is pervasive or whether it arose through some quirk of the experimental procedure.

With regard to the fairness observed in the DLUG, one possible explanation is that it is due to uncertainty concerning the actual work required. Recall that participants were asked to complete the DLUG before being told anything about the search topic – our motivation for this was to explore DLUG behaviour in terms of notional work units stripped of any particular attractiveness or unattractiveness. Since the chances of this occurring in the real world are probably quite low, a more realistic design would provide absolute task knowledge to participants. An additional problem with the present study is that, while the preference towards fair offers does at least appear stable, the number of participants is clearly too low to establish the veracity of the findings. To address these issues, we designed an experiment that builds on our first study. In the second experiment, we attempt to directly manipulate qualitative aspects of the search task by informing participants of their search topic in advance of the DLUG, manipulating topic as an independent variable. This allows us to explore how information search topic might influence both allocation agreements and strategies that emerge during the process of completing work. We also obtain further evidence concerning the apparent matching of task completion times. This study is reported in the following section.

5.3.7 Study 2 Summary

This study was created in order to gain an initial grasp on behaviour in our division of labour ultimatum game. Through a study involving 20 participants, we found that:

- The modal allocation of was an even split, occuring in eight of 10 cases.
- Two methods were used to coordinate work. In the first, participants simply waited until duplications arose and then repaired them on-the-fly. In the second, participants explicitly agreed on a partitioning strategy that saw them each take responsibility for segregated aspects of the available Web or document space.
- Participants' completion times were synchronised in an interesting yet unexpected fashion.

5.4 Thesis Study 3: The Impact of Search Topic in the DLUG

5.4.1 Study Overview

Following on from the experiment reported in Section 5.3, the study reported here was designed to explore the veracity of our prior findings. Our goal was to examine three salient issues:

- In experiment 1, we saw that the modal allocation of work in our DLUG was a fair split, mirroring findings from the standard UG. Our model was intended as a pure instantiation of the DLUG, but was somewhat unrealistic in that participants lacked knowledge about the task at hand. Thus, we sought to rectify this concern by providing complete task knowledge up-front to participants.

- As well as exploring how information search topic might influence allocation agreements, we wanted to explore the impact of topic on the coordination strategies used by our participants, and to further explore the prevalence of different coordination strategies as a function of the task at hand.

- Finally, we wanted to examine whether the search time effect would be replicated in an experiment with more participants.

We elected to address these three issues simultaneously by manipulating search topic as an independent variable. In our first study, we found that the information topic of ‘design psychology’ was not familiar to most of our participants. Previous work has shown that knowledge of a topic can aid Web search, whereby searchers can use existing knowledge to form more effective queries, in turn leading to speedier information retrieval (Duggan & Payne, 2008; Kang *et al*., 2010). Beyond this, we hypothesised that knowledge of a topic might also allow new strategies for partitioning the joint task among team members. The study therefore attacks the issue of topic knowledge and its dual impact on division of labour *and* resultant coordination during collaborative information seeking.

5.4.2 Design

The design of this study was similar to study 2 in that we retained the scenario of dividing a task requiring 10 bibliographic references, presented as forming a reading list for a night class. Again, participants worked anonymously and were located in separate rooms. We used two different search topics in the study, manipulated between groups and initially forming two experimental conditions. Our choice of topics aimed to provoke differing responses from our participants while balancing feasibility, i.e., sources on the topic had to actually exist so that participants were able to do the work requested.[20] The topics we elected to use were as follows:

[20] We recognise that it is likely impossible to find a research topic that is universally boring or interesting, if only because of interpersonal differences in interests. Since one could potentially spend forever trying to find a topic that is universally unattractive, we eventually settled on ‘good enough’ topics that provoked a sufficiently quantifiable difference in terms of interest and enjoyment from our participants.

1. *Popular music in the 20th century*. This topic was selected because most people have at least some passing knowledge of musical acts, and might even have personal preferences that they could draw on during the task, meaning that popular music would be a relatively attractive, or at least non-aversive, topic for creating a bibliography. A straightforward search for this topic on Google returned more than 37,700,000 hits, suggesting a sufficient number of sources would be available for our task.

2. *The life cycle of mycetozoa slime mold*, which we assumed most participants would be ignorant of and uninterested in. This topic was selected based on the relative obscurity of slime molds and because participants in pilot trials indicated that the topic appeared daunting at first glance. The topic was also manageable in terms of available sources, returning more than 61,600 hits on Google.

Pilot studies confirmed our intuitions about the attractiveness of these topics (using the scales described in the Materials subsection below) and indicated that both tasks were manageable within our defined experimental constraints ($<$1 hour per trial) without the task being trivial.

5.4.3 Materials

All locations, equipment, and software were identical to those used in study one. The procedure and experimental protocols were the same as before, with a small modification to the instructions such that the search topics were presented up-front alongside the DLUG. (Materials can be found in Appendix D.) As in study one, we explored participants' judgements of topic familiarity via a 5-point Likert scale included in the post-experiment questionnaire. To probe differences between our two conditions, we added three 5-point scales to explore participants' judgements of topic difficulty (i.e., how hard it was to obtain results), interest in the topic, and overall enjoyment of the task.[21]

The experimental task was presented to participants as follows: *You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about [artists or bands that shaped the history of popular music in the 20th century] / [the life cycle of mycetozoa slime molds].*

In this study, a non-intrusive Firefox extension, *HCI Browser* (Capra, 2010), was used to record search behaviour surreptitiously, including queries and page visits, into a text file.

5.4.4 Participants

Thirteen male and 27 female students, research and administrative staff from 12 different University departments took part in this study. Participants' ages ranged from 18–48 (*M* = 25.7). Again, we did not control for demographic variables and participants were randomly assigned to condition and role.

[21] We note that it would be preferable to examine ratings of topic interest both before and after the study, so as to allow for pre and post-hoc comparisons. However, we decided that administering a pre-task questionnaire, in the gap after the DLUG but before commencing the task, would be impractical from a pragmatic perspective and would also draw participants' attention to topic as a matter of interest.

There were 10 pairs in each condition; in the popular music condition, one pair was male-male, four were female-female and the rest were mixed pairings. In the slime mold condition, one pair was male-male, five were female-female and the rest were mixed pairings. All participants were naive to our experimental design and were not involved with our previous study. We again evaluated computing skill and search experience using measures from Morris (2008). All participants reported conducting Web searches daily. Again, our participants were comfortable with Web search and were not novice computer users, with 11 participants reporting expert computing skill and 29 reporting intermediate. Participants were recruited via the University noticeboard on the promise of £5 for completion of the experimental task.

5.4.5 Procedure

The procedure employed in this experiment was identical to that used in study 2, see subsection 5.3.4.

5.4.6 Results

5.4.6.1 Manipulation Checks

Participants in the slime mold condition rated themselves as significantly less familiar with the search topic (*M* = 1.25, *SD* = .786) than those in the pop music condition (*M* = 3.35, *SD* = 1.348), t[38] = -6.016, $p < 0.001$. Those researching slime mold also considered it significantly more difficult to find sources (*M* = 3.15, *SD* = 1.266) than those researching pop music (*M* = 2.30, *SD* = .865), t[38] = 2.534, p = 0.016.

Participants in the slime mold condition considered the topic significantly less interesting (*M* = 2.25, *SD* = .1.251) than those in the pop music condition (*M* = 4.05, *SD* = 1.050), t[38] = 4.928, $p < 0.001$. Their rated enjoyment of the task was not significantly different between conditions.

5.4.6.2 Workload Allocations

Table 5.2 shows the allocation data for the pop music condition, and Table 5.3 displays the slime mold condition. It can be seen that no proposals were rejected. Regarding allocations, eighteen of the twenty proposals were 5-5. The two unfair splits were made at 6-4, and both of these occurred in the slime mold condition. Notably, both involved *proposers* offering to take on more work. Unlike in experiment one, neither of these two allocations were corrected during the work process. Instead, the pairs completed their allocations (6-4) as agreed.

Session ID	Proposer Allocation	Responder Allocation	Acceptance
2.1	5	5	A
2.2	5	5	A
2.3	5	5	A
2.4	5	5	A
2.5	5	5	A
2.6	5	5	A
2.7	5	5	A
2.8	5	5	A
2.9	5	5	A
2.10	5	5	A

Table 5.2: Workload allocations and acceptances, study 3, pop music condition. *A = Accepted, R = Rejected*.

Session ID	Proposer Allocation	Responder Allocation	Acceptance
2.11	5	5	A
2.12	6	4	A
2.13	5	5	A
2.14	5	5	A
2.15	5	5	A
2.16	6	4	A
2.17	5	5	A
2.18	5	5	A
2.19	5	5	A
2.20	5	5	A

Table 5.3: Workload allocations and acceptances, study 3, slime mold condition. *A = Accepted, R = Rejected*.

5.4.6.3 Motivation for Dividing Work

In the pop music condition, nine proposers cited fairness as their primary consideration when deciding how to allocate the work. Two proposers stated that the fear of rejection also played a role in their decision. One proposer used his own level of expertise as a basis for making a fair split, anticipating that his partner could cover the areas of music where he had little knowledge.

In the slime mold condition, seven proposers invoked fairness when choosing an allocation. One of these considered a fair split as the most efficient way to divide the work. Three proposers stated that their proposal came about as a method of avoiding rejection. The two proposers who took on six sources for themselves both rationalised this in terms of an assumption that they were likely to be better at the task than their unknown partner.

5.4.6.4 Coordination Strategies

Table 5.8 shows the strategies identified in this experiment together with the number of pairs who used each, delineated by search topic. As before, *Wait & Repair* was the most frequent strategy, but in only one case was repair actually necessary. As observed in experiment one, three pairs employed the *Partition Web/ Document Space* strategy. The vignette below illustrates one pair explicitly coordinating work by partitioning only the Web space:

> [Pair 2.19, Slime Mold]
> Participant One: shall we using 2 different search engine to find the sources? or else we may have same results
> Participant Two: no problem... which one are you using?
> Participant One: i am using google how about you?
> Participant Two: i am using google scholar... will also look into library resources
> Participant One: maybe you can mainly focus on the scholar and i will search for the relevant websites
> Participant Two: ok

We also identified a new strategy, which we call *Partition Semantic Space*. Use of this strategy was confined to the pop music condition, and involved choosing different aspects of the topic independently of web services or document types and dividing these aspects between collaborators. Five pairs explicitly discussed and employed this strategy. The vignettes below illustrate two different pairs using this strategy to coordinate after agreeing 5-5 allocations:

> [Pair 1.1, Pop Music]
> Participant One: would you like to spend one minute discussing how to carry out the task?
> Participant Two: ok
> Participant One: would you like to take a particular historical period?
> Participant Two: does the 20th separate nicely?
> Participant One: well I was thinking, before rock and roll and after rock and roll... for example: before 1950s... and after 1950s when all the rock and pop music starts
> Participant Two: sounds good... although there was more after 1950
> Participant One: so if you are comfortable with it I am happy to do the first 'half' of the X century, say the beginning of popular music, early blues and ragtime... and you can do the 1960s onward... so we are sure we don't replicate
> Participant Two: sounds like a plan

> [Pair 1.8, Pop Music]
> Participant Two: so do you want to take Hendrix and one of the others and I'll do the other two
> Participant One: yes... I will take hendrix and Micheal [Jackson] pls
> Participant One: ok, ill do the beatles and Elvis

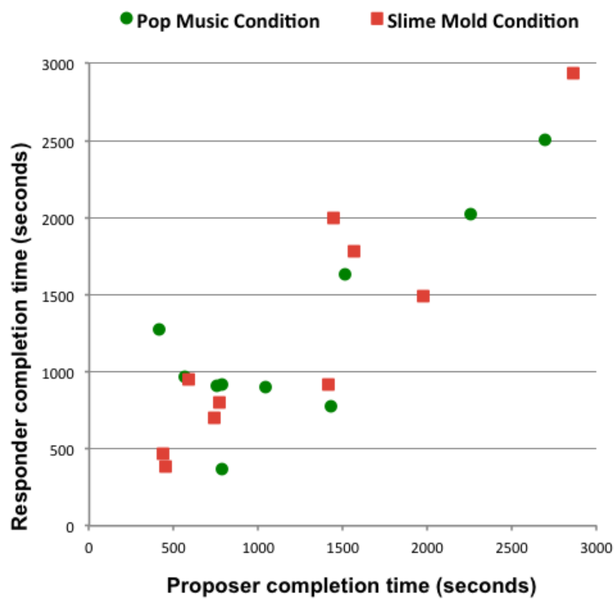

Figure 5.2: Correlation between task completion times of proposers and responders, study 3, r_s = .58.

5.4.6.5 Task Completion Times

Figure 5.2 graphs the raw search time data and again shows a striking concordance between partners' search times. Again, a Pearson correlation was computed on log transformed data, and revealed a strong positive correlation, $r(18) = 0.76$, $p < 0.001$.

5.4.7 Discussion

Our manipulation of the attractiveness of the work task, in terms of topic of search, was successful. Participants considered the slime mold topic significantly less interesting and significantly more difficult to obtain sources for. They also considered themselves significantly less knowledgeable about the topic. However, this manipulation had no observable effect on the outcome of the DLUG. Again, the majority of offers (90%) were 5-5 and all offers were accepted. Although there were two 6-4 offers in the slime mold condition, the numbers involved are not large enough for this to be any more than a hint of a manipulation effect. That said, this hint is elaborated slightly by these two proposers explaining their offers in terms of their own ability to perform the task, which is conceivably influenced by the perceived difficulty of the search topic. This rationale also gives us some insight into the issues that proposers take into account when deciding upon their allocation, directly implying that division of work is less straightforward than division of money. (This subject is discussed in more detail later in this thesis).

Unlike experiment one, neither of the two uneven allocations was altered in the process of work. Looking into the process data, the likely reason for this is that the participant with more work managed to finish their part of the task first. It is interesting that these inequitable offers do not appear to have been motivated by disparity in knowledge; the individuals in question both considered themselves very unfamiliar with the slime mold topic. Instead, it appears that allocations were made on the basis of perceived skill, i.e. that the proposer was sufficiently confident in their own abilities that they kept more work to ensure that it was completed quickly. As mentioned above, it appears that these perceptions were well-justified.

The strategies of Wait & Repair and Partition Web/Document Space were observed again in the slime mold and the pop music conditions, where additionally a new strategy, Partition Semantic Space, was observed in five out of ten pairs. In this case, participants leveraged existing knowledge to establish coordination. We suggest that this strategy depends on both participants having some pre-existing knowledge of the topic in question, and expecting their partners to share this knowledge. But, when a topic is obscure or unfamiliar, such behaviour is not possible, necessitating the use of alternate strategies (such as partitioning of the Web space) to create coordination.

Our third concern in this experiment was the replication of the search time effect obtained in study 2 (see Section 5.3 above). Although the correlation was slightly weaker in this experiment, we did observe the same overall trend, indicating that it was not a fluke result. Observation of Figure 5.2 shows the effect to be quite striking. We can see that there are several trials which ended quickly, lasting a little under 500 seconds (8 minutes). At the other extreme, we see trials that lasted almost 3000 (50 minutes). Despite a small amount of variance within some pairs, completion times appear to be highly similar within each trial. This again seems to indicate a desire for overall equity in both the allocation and completion of work. However, at this point we have no evidence beyond this speculation as to what could drive this effect, what could regulate it, and how it could be bootstrapped. Informal remarks by participants suggested that some felt 'pressured' to increase their own work rate in order to keep up with their partner. This suggests that asking participants for more information about issues concerning their regulation of effort and the attention they afford to their partner could be instructive in terms of explaining the effect.

In general, the evidence we have gathered thus far from our DLUG is indicative of fairness preferences in the division of labour. However, it is important to note that proposers in the UG often behave strategically, making fair offers in order to guarantee acceptance from the responder (Guth & Tietz, 1988; Ochs & Roth, 1989; Kravitz & Gunto, 1992). This was an important consideration for at least two proposers in experiment one, and a further five in experiment two, all of whom stated that they had wanted to make an unfair offer but chose not to do so. Given that proposers seem interested in lowering their workload while avoiding rejection, a sensible offer in the DLUG is a proposal that provides the best possible chance of avoiding rejection whilst still lowering the workload by as much as possible. This line of reasoning, combining gain in utility with probability of rejection, may explain the frequency of fair offers in our study. It is therefore possible that, just as in the classic UG, fair offers arise as a result of strategic behaviour intended to mitigate risk, rather than a desire for fairness *per se*.

In Chapter 4 we saw that a classic way of investigating strategic fairness is via the dictator game, where the responder's option to reject the allocation is removed. The proposer simply allocates money, and the responder takes whatever he is given. The supposition here is that if proposers are truly concerned about equity and fairness then they should continue to divide the surplus in half even when there is no threat of rejection. The canonical model predicts that proposers in dictator games should never give any money away. While studies typically show that participants do continue to make positive allocations in dictator games, one recent meta-analysis of dictator games found that the average offer was 28.35% of the pie (Engel, 2011). When this figure is compared to the average 40% offer in ultimatum games (Camerer, 2003), the evidence gives credence to the idea that proposers play fair (or stay relatively close to fairness) due to fear of rejection in ultimatum games.

To follow up on these issues, we designed a third study that explores whether threat of rejection could explain the prevalence of fair offers. The benefit of our DLUG is that it lends itself directly to a dictator style design—we can simply remove the responder's opportunity to reject and explore whether proposers continue to allocate half of the work to their anonymous counterpart. Such a design is also interesting for the procedural components of our results, in that we may be able to examine the persistence of the matching effect under circumstances of unequal workloads.

5.4.8 Study 3 Summary

This study was created in order to explore the impact of information search topic on behaviour in our division of labour ultimatum game. Through a study involving 40 participants, we found that:

- The modal allocation was once again an even split, occuring in 18 of 20 cases. No offers were rejected.
- A new strategy for coordinating work emerged in the condition with popular music as the topic of interest. This strategy, Partition Semantic Space, complements our earlier findings from study 2, and demonstrates another emergent method of preventing redundancy during collaborative search.
- We replicated our earlier finding concerning an apparent matching of participant's task completion times.

5.5 Thesis Study 4: Testing the Persistence of Equity in a Dictator Variant

5.5.1 Study Overview

In study 3 we found that the attractiveness of the work task was not enough to steer allocations away from the point of fairness. Although the results provide further evidence of a fairness norm in the division of labour, one issue the study did not address was that of fear of rejection may play a role in creating fair offers. This rationale was uncovered in our reading of post-experiment questionnaires, and has been used by economic theorists to explain behaviour in ultimatum experiments (cf. Guth & Tietz, 1988; Ochs & Roth, 1989; Kravitz & Gunto, 1992).

The present study draws upon the literature on dictator games, reviewed in Chapter 5 of this thesis, to explore threat of rejection as an explanation of our results. We achieve this by removing the responder's opportunity for recourse from the bargaining procedure. Proposers are then given 10 work items and told to allocate any number to the responder.[22] This means that if proposers are truly concerned about equity and fairness, egalitarian offers should persist in a setting where there is absolutely no threat of punishment. If, however, fair offers arise through strategy, we should see a drop in eql allocations and a shift towards selfish behaviour. Proposers are thus confronted with a very stark choice between sharing at least some of the workload or free-riding entirely.

The canonical prediction in a dictator game is for the proposer to keep everything and for the responder to receive nothing. As with our DLUG, the reverse is now true: if the proposer is solely interested in receiving the £5 for attending the experiment, and is given a costless choice between completing more or less work to receive that reward, he should prefer less work. In a situation where he can delegate as much work as he chooses with no risk of punishment, the extreme prediction would be a 0-10 allocation, with all of the work going to the responder.[23] The proposer would then leave the experiment with his or her money, and the responder would complete the work as assigned. However, the dictator game literature tells us that it is usually rare for such allocations to occur unless they are somehow legitimised, e.g. via an earnings task (Hoffman *et al*., 1994; Cherry, 2001; Cherry *et al*., 2002; Oxoby & Spraggon, 2006; List & Cherry, 2008). That said, offers in dictator games do tend to be smaller than those in ultimatum games, meaning that results typically fall somewhere between the markers of fairness and pure self-interest. Our intuition, then, is that while the 0-10 allocation may be too extreme an outcome, we may at least see a shift away from fair offers when responders cannot reject allocations.

In addition to exploring the delegation scenario outlined above, we wanted to explore whether an exemplary manipulation of the UG would apply when replicated in the context of division of labour. We saw in Section 4.3 of Chapter 4 that the literature on economic games is quite vast, and a true exploration of the many possible variables is beyond the scope of a single thesis. For the present experiment, we opted

[22] The second player has no opportunity to respond so is not really a 'responder' as such. We continue to use the word responder for continuity and to distinguish between player roles.

[23] As noted in Chapter 5, dictator games do not typically include a minimum offer because the zero allocation is meaningful. We follow similar reasoning and allow proposers to delegate all of the surplus.

to impose an information asymmetry an manipulate *available information* as a variable of interest. This allows us to further address the question of whether or not proposers actually want to be fair, albeit in a different way to that of the dictator game.

In Subsection 4.3.2.7 we saw that the information held by each player in an ultimatum game can affect outcomes, particularly when this information is asymmetric. For example, when proposers know the amount to be divided but responders do not, proposers typically lower their offer to increase their payoff slightly (Croson, 1996). Responders are also more likely to accept such offers, since they cannot determine the relative fairness of the allocation (Croson, 1996). Extending this logic to circumstances of division of labour, let us imagine that an information asymmetry exists between players, where the proposer knows the amount to be divided but the responder does not. In such a situation, the proposer could allocate *more* work to the responder if the responder does not know the exact amount to be divided. For the purposes of dictation, information asymmetry may make it sufficiently more easy for a proposer to be unfair if the responder is not initially aware of the total amount to be divided. Even though the responder has no opportunity to reject, he or she has no way of knowing that the allocation is secretly unfair, and thus collaboration could proceed in an otherwise typical fashion—he or she has simply been 'dictated to' and there is no way of knowing whether this dictation is fair or unfair.

The following experiment sought to explore the established intuitions about information asymmetry, but under circumstances of division of labour. If proposers care about fairness, the asymmetric information manipulation should have no effect on offers. On the other hand, the proposer might be tempted to take advantage of the responder and delegate more work in order to leave the experiment early.

We also considered that it would be interesting to explore the impact of inequity in the planning stage of work, in terms of how it could affect coordination strategies and the search time effect reported earlier. With regard to the latter, two possibilities exist. First, if highly unfair allocations were made, we might expect this to disrupt the matching effect. Alternatively, it would be interesting to explore whether the participant with a greater workload upped their effort in order to leave the experiment around a similar time as the anonymous partner. It would also be interesting to see how the matching effect plays out when one person does not have full knowledge of the collective goal, as would be the case under circumstances of incomplete information about the task at hand.

Lastly, we used this experiment as an opportunity to probe the causes of the search time effect. In study 3, some participants had stated that they felt pressured by their partner's work speed. We therefore augmented our post-experiment questionnaire with questions that tried to gauge whether participants felt influenced by their partner's activities; whether they monitored the activities of their partner during the study; and whether they were satisfied with the speed at which their partner worked. These questions are described more precisely in our Materials subsection below.

5.5.2 Design

The design of this study was highly similar to that of studies 2 and 3 in that we retained the bibliography task for consistency purposes. This experiment used two conditions. First, a standard dictator game, identical in execution to our DLUG but with the sole caveat that responders could not reject the allocation. In this condition, both participants knew the total amount of work to be divided, i.e., responders were told that there were 10 items. Our second condition created an information asymmetry, whereby, at the time of task delegation, responders were not told the amount to be divided, and proposers were aware of this fact. We refer to these conditions as complete information (CI) and asymmetric information (AI) respectively. We elected to retain the slime mold topic from study 3, using it in both conditions. Holding this constant was necessary to ensure that any changes in offer distribution (as compared with studies 2 and 3) could be correctly attributed to the removal of rejection from the bargaining scenario, rather than a change in search topic.

5.5.3 Materials

All equipment and software were identical to those used in studies 2 and 3. Participants again worked anonymously using an MSN client for communication (shown in their left monitor) and a Firefox Web browser to find information (right monitor). Experimental protocols were similar to those used in studies 2 and 3, but were modified to account for changes to the allocation procedure. (Materials can be found in Appendix E.) In the CI condition, proposers were told they could allocate any number of sources to the responder. In the AI condition, proposers received the same instructions with an additional paragraph explaining that the responder did not know the amount to be divided. Instructions given to responders were the same in both conditions, except that in the AI condition responders were not told the exact number of items required in the reading list. Instead they were told that the proposer would allocate ‘a number of sources’ to them. After pilot trials it emerged that proposers were unsure whether a completely selfish allocation would affect their payment. Proposers were also unsure whether 0-10 was a legitimate allocation for the purposes of the study. We therefore modified the instructions to clarify these issues.

In this study we used different laboratories to studies 2 and 3; this came about as a result of our department’s relocation to another area of the university campus. Despite the change in physical location, we attempted to replicate the pragmatics as closely as possible. The laboratories used after the move were at opposite ends of a communal research lab with about 100 yards’ separation. These labs were accessed via separate doors, meaning that participants were not visible to one another. The experimenter did his utmost to preserve anonymity during the study and participants did not meet in accordance with our established experimental protocol. Proposers and responders were told to meet the experimenter in different areas of the building, meaning that they did not see or meet each other (as would occur if asking them to come directly to the lab).

5.5.4 Participants

Sixteen male and 24 female students, research, administrative and support staff from 18 different University departments participated in this study. Participants' ages ranged from 18 – 38 years (*M* = 24.7) making the sample comparable to earlier experiments. Participants were recruited via the university noticeboard on the promise of £5 for participating in a research study. We did not control for demographic variables and participants were randomly assigned to condition and role.

There were 10 pairs in each condition; in the CI condition, one pair was male-male, two pairs were female-female, and the rest were mixed pairings. In the AI condition, one pair was male-male, four pairs were female-female, with the rest mixed pairs. All participants were naive to our experimental design and had not participated in our previous experiments. We again evaluated perceived computing skill and search frequency using measures from Morris (2008). No participants were novice computer users and all reported conducting web searches at least several times per week.

As in study 3, we gauged participants' perceptions about topic familiarity, difficulty and interest. The topic of slime mold was not familiar to most of our participants (*M* = 1.4, *SD* = 0.9), was considered moderately difficult (*M* = 2.95, *SD* = 1.1), and was not particularly interesting (*M* = 2.3, *SD* = 1.039). These results are highly similar to those recorded in the slime mold condition of study 3.

5.5.5 Procedure

The procedure employed in this experiment was identical to that used in study 2, see subsection 5.3.4.

5.5.6 Results

For clarity we again present our results in sequence, focusing first on workload allocations, reported motivations for dividing work, coordination strategies employed to prevent redundancy, and the search time effect.

5.5.6.1 Workload Allocations

Tables 5.4 and 5.5 show the breakdown of allocations in each condition. It can be seen that, as in the DLUG, the modal outcome was a fair split of the work. Eight of 10 allocations were 5-5 in both the CI and AI conditions. In the CI condition there were two 4-6 allocations; in the AI condition, there was one 4-6 allocation and one 3-7 allocation.

Session ID	Proposer's Workload	Responder's Workload
3.1	4	6
3.2	5	5
3.3	5	5
3.4	5	5
3.5	5	5
3.6	5	5
3.7	5	5
3.8	5	5
3.9	4	6
3.10	5	5

Table 5.4: Workload allocations, study 4, complete information (CI) condition.

Session ID	Proposer's Workload	Responder's Workload
3.11	5	5
3.12	5	5
3.13	3	7
3.14	5	5
3.15	5	5
3.16	5	5
3.17	5	5
3.18	4	6
3.19	5	5
3.10	5	5

Table 5.5: Workload allocations, study 4, asymmetric information (AI) condition.

5.5.6.2 Motivation for Dividing Work

In the CI condition, 5 proposers mentioned fairness when explaining their allocation in the post experiment questionnaire, e.g., *"I think allocating half work to each is the fairest option."* [ID 3.5, see Tables 5.4 and 5.5]. Not surprisingly, all of these proposers made 5-5 allocations. Two other proposers mentioned time and efficiency when explaining their offers, e.g., *"Chose to distribute the work evenly as it seemed like the fastest way to get the task done."* [ID 3.10]. The remaining 5-5 allocator's rationale hints at fairness without mentioning it explicitly: *"50-50 dividing of the work seemed like the reasonable thing to do, regardless of the task context."* [ID 3.6].

The questionnaire responses provide some insight into the decisions of the 4-6 allocators in the CI condition. The first stated that he was tempted by a highly unfair allocation: *"Given a free choice between 0 and 10 sources I was tempted by 10, however I chose to search for 4; easier task for me"* [ID 3.1]. The other participant's rationale was slightly unclear but appears to hint at the perception of ability: *"Just want to know/try to find sources by myself"* [ID 3.9].

In the AI condition, six of 10 proposers cited fairness as an explanation of their allocation. All of these proposers made 5-5 allocations. Two other proposers mentioned equality, but not fairness: *"2 participants: equal distribution of workload. This is the fastest way to complete."*. [ID 3.12]. *"I chose to tell my partner to find 5 of the sources so as to make the work equally difficult."* [ID 3.19]. The former quotation again mentions speed as a factor of concern when allocating work.

The rationales given by the two unfair allocators were as follows: *"I gave more work to the other participant because I didnt want to do much by myself."* [ID 3.13] *"I gave my partner 6 and me 4. Mainly because I don't feel very confident using the internet search engine for educational topics."* [3.18].

5.5.6.3 Coordination Strategies

Fourteen of our 20 pairs employed the Wait & Repair strategy. Two pairs incurred a breakdown resulting from duplication, and again in both cases, the person responsible for the duplication went on to correct the mistake and obtain another source:

> [Pair 3.16, AI condition]
> *[Participant one pastes a reference]*
> Participant Two: I've already used the creationwiki source
> Participant One: I will refind one

The remaining six pairs employed the Partition Web / Document space strategy by explicitly discussing their strategy, as illustrated by the following vignette:

> [Pair 3.5, CI condition]
> Participant Two: Hi there... I am going to have a look on JSTOR first of all for journals
> Participant One: Hi, sure – I'll look on Google Scholar and GoodReads

In one case, partitioning of the web space was conflated with the Wait & Repair strategy:

> [Pair 3.6, CI condition]
> Participant One: I'm guessing the best way to go about it would be to either look into books which are related to the subject, or papers
> Participant Two: yes, we might might be able to find good sources from wikipedia
> Participant One: Should we use the same sources... or different ones? Using the same might imply identical readings hence giving us more work... what do you think?
> Participant Two: no, we have to find unique sources but we could find them from the same webpages
> Participant One: True. We'll just check the links once we post them here and see whether they're the same

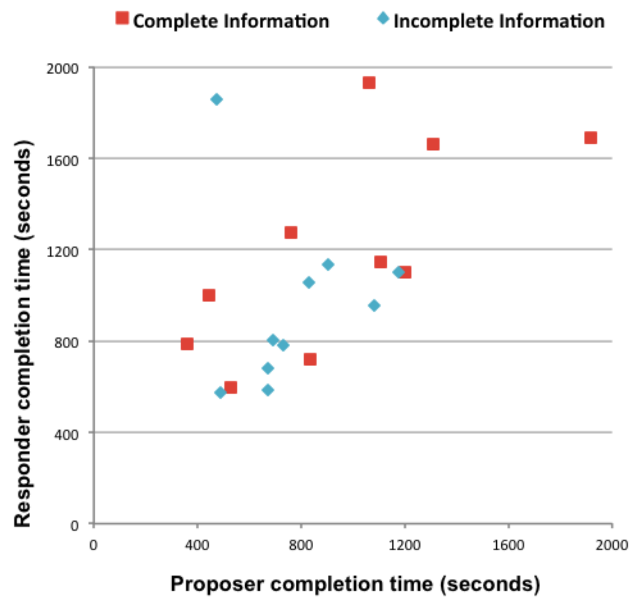

Figure 5.3: Correlation between task completion times of proposers and responders, study 4.

5.5.6.4 Task Completion Times

Figure 5.3 graphs the raw search time data and appears to show the same trend between paired search times seen in studies 2 and 3. Again, the data was positively skewed (skewness statistic = 0.96) so the data was transformed using natural logarithm. A Pearson's product-moment correlation was computed and revealed a moderate correlation, $r(18) = .513$, $p = 0.021$. Although this correlation is somewhat weaker than the trend observed in experiments one and two (where both r coefficients > 0.75), we can see that Figure 5.3 does appear to display the same overall trend as seen in our prior experiments.

Appraisal of Figure 5.3 suggests one potentially outlying data point (upper left region of the chart) which may be responsible for lowering the strength of the correlation. Such an outlier could be tolerated in a large sample but it is likely that this pair, who appear to have quite a large difference in their completion times, significantly weakens the correlation in our relatively small dataset. We explored the impact of removing this data point in line with literature on outlier removal (Judd & McLelland, 1989; Barnett & Lewis, 1994; Osborne & Overbay, 2004), which suggests that outliers with a z score over 3 should be removed. Exploratory analysis revealed that pair 3.11 from the AI condition (see Table 5.5) were eligible for removal based on this criterion, where z scores were computed based on the mean difference in completion times between each pair ($M = 282$ seconds difference, $SD = 339$). With a difference of 1382 seconds, pair 3.11 were more than 3 standard deviations from the mean; temporary exclusion of their data from the analysis results in a positive correlation closer to the expected strength, $r(17) = .704$, $p < 0.001$, Pearson's product-moment coefficient.

Although exclusion of this pair does return the correlation to the anticipated level, we do not argue

Category	Description	Frequency	Example Statement(s)
Workrate Adjustment	Makes statement about adjusting their workrate based on that of their partner, i.e. indicating that they either sped up or slowed down.	15	*"Because that person finished his part so fast, I think I have to hurry up and finish it soon."* [R, ID 3.3] *"Felt I had to find sources as quickly!"* [R, ID 2.13]
Coordination of Work	Makes statement about adjusting their search behaviour, i.e. altering their choice of pages visited or search engines used.	8	*"I became more abstract or alternative in my wording and what I looked for"* [R, ID 3.5] *"I would find alternative source on the same website"* [P, ID 3.14]
Ensuring Quality	Makes statement about adjusting the way in which they judge the quality of found sources.	2	*"He/she makes me feel like I should check his/her sources (how reliable)"* [P, ID 3.9]
Miscellaneous/Other	Makes statement with content that does not match other categories or statements.	4	*"We both find the relevant sources one by one, like inspiration"*[R, ID 3.6]

Table 5.6: Participants' responses to the question "Did you feel influenced by the activities of your partner during the study?", study 4. In Example Statement(s) column, P = proposer, R = responder.

that they should be discarded outright. Post-hoc examination of the pair's chat log, questionnaires and search history implies that the disparity may have arisen due to a lack of advance planning about where to search, with the second participant stating that *"his choices* [referring to the proposer] *made me change mine"*. In other words, failure to coordinate perhaps resulted in both persons searching similar material, with the responder repeatedly pipped to the post in her choices. This would in turn necessitate further search, perhaps explaining the disparity in completion times.

5.5.6.5 Search Time Effect: Questionnaire Responses

In order to probe the causes of the search time effect, participants were asked whether they felt influenced by the activities of their partner during the experiment. Of our 40 participants, 38 responded to this question. The majority (29) stated that they did feel influenced in some way, with only 9 stating 'no'. The remaining two responses were unclear or nonsensical.

Since we were interested in whether participants felt compelled to alter their rate of work in order to match their partner, rather than how participants selected sources for the bibliography, we classified affirmative answers using an open coding approach. This led to the formation of four categories of response, as shown in Table 5.6. Responses were grouped into these categories by two independent coders; initial agreement was .93 with two disagreements later resolved through discussion. Table 5.6 shows that statements pertaining to the adjustment of work rate, based on the feeling of being influenced the partner's work, were the most frequent. Such statements typically described feeling the need to either speed up or slow down the pasting of references in order to 'keep pace' with the partner.

5.5.7 Discussion

The present study was motivated by the finding that some participants in studies 2 and 3 explained their fair offers as a strategic method of avoiding rejection in the DLUG. Here, we used an applied version of the dictator game to investigate whether proposers would continue to allocate work 'fairly' when the responder could not reject the allocation. Our intuition was that, if fair allocations in prior experiments were the result of strategic behaviour rather than fairness preferences, allocations might shift towards unfairness when proposers are able to allocate work with impunity. To further probe this intuition, we implented a condition involving asymmetric information, where only the proposer had complete information about the total workload to be divided. This manipulation removes the responder's ability to compare his payoff to the proposer's; such a design has been shown to suppress fair allocations in both ultimatum and dictator experiments (Croson, 1996).

In spite of our design, 16 of 20 proposers made a 5-5 allocation. Moreover, our implementation of information asymmetry appears to have had almost no effect on allocation behaviour. Both of these findings stand in stark contrast to prior results from the economic literature. Regarding unfair splits, three were made at 4-6 and one was 3-7, and all were in the proposer's favour, i.e. proposers delegated more work to responders. This is slightly different to our previous experiments in that earlier allocations typically saw proposers keep more work for themselves, perhaps suggesting that the dictator manipulation had a small impact. However, the majority of participants still opted for an even split. Our decision to explore the dictator variant was partly motivated by the intuition that fair offers may arise through a mixture of strategy and other-regarding preferences. We hinted in our discussion of study 3 that people may seek to minimise risk by making an equitable offer—such an allocation lowers the threat of rejection while also providing a reasonable reduction in workload. The fact that people persist with equitable offers when there can be no rejection implies that risk minimisation cannot account for our findings. Instead, other motivations must underly the tendency towards fairness.

We have now observed that a fair split was the modal outcome in three studies. Of 50 recorded splits, 80% have been equitable and the remainder stray from equity by only a small amount. Face value appraisal of our results suggests that fairness norms in division of labour may be stronger than those for monetary payoffs. However, we must be careful about this conclusion given the presence of several other competing explanations. For example, examination of the post-experiment questionnaires revealed that some dictators had wanted to be unfair but chose not to, with one dictator stating: *"Given a free choice between 0 and 10 sources I was tempted by 10, however I chose to search for 4"*. Such responses are interesting in the sense that they demonstrate participants reversing on their initial preferences. However, they provide no clear insight as to *why* this reversal occurs. One very distinct possibility is that the moral cost associated with the highly unfair action, in this case a 0-10 allocation, outweighs the gain in utility associated lowering the workload. Recent models of individual decision making (e.g. Levitt & List, 2007) suggest that utility functions may incorporate a nonpecuniary moral payoff, denoted by the negative externality that a selfish decision imposes on others. Here, the decision to engage in the 0-10 allocation may be at odds with an individual's identity (Akerlof & Kranton, 2000) or may be viewed as immoral, antisocial, or simply wrong (cf. Levitt & List, 2007). This would explain why participants chose to renege on their initial tendencies, opting instead for fair splits. One other possibility is that participants chose not to follow through on the unfair allocation due to the potential for scrutiny from the experimenter. That is, participants may have been aware that a reduced workload would result in early departure from the experiment, in turn signalling a highly selfish allocation to the facilitator. Such scrutiny may have deterred participants due to fear of looking like 'the bad guy'. However, it would be difficult to deal with this problem without implementing a convoluted double-blind procedure. These concerns hint at deeper underlying issues with the enactment of economic experiments more generally. For now, we shall defer in-depth discussion of these possibilities until a later, more general discussion of the present experiments (section 5.7 of this chapter).

Turning to our process results, we again observed the use of strategies to coordinate work during collaborative search. Although no new strategies emerged, we were able to replicate the strategies of Wait & Repair and Partition Web/Document space, as observed in studies 2 and 3. Additionally, we did not observe the use of the Partition Semantic Space strategy. Since participants reported low knowledge about the subject, these findings support our arguments that semantic partitioning is only available when at least one collaborator has at least some knowledge of the topic.

This experiment also replicated the search time effect for a second time. While the strength of the correlation was initially moderate, we saw that removal of one outlier from our dataset returned the correlation to the expected strength. This again suggests a desire for fairness in the completion, as well as allocation, of the collaborative task. Recall, however, that the effect has two components: first, the matching of completion times *within* a pair, and second, an often-considerable variance *between* pairs. For example, in the present experiment we see pairs with completion times of around 10 minutes, yet others whose time approaches 30 minutes. This begs the question as to what exactly determines the overall task completion time for the pair—it is not immediately clear as to why one pair would complete the task quickly when another takes three times as long.

One speculation that arose from study 3 was that participants somehow felt compelled to raise or lower their workrate based on the observed efforts of their partner. Through our analysis of post-experiment questionnaires in the present study, we found some that participants reported a perceived need to either speed up or slow down the pasting of references in order to 'keep pace' with the partner. This would explain how participants finish up with similar times. However, none of the statements tell us exactly *why* participants felt the need to do this. It may be that the effect relates to some overall desire for equity in both the allocation and the process. Alternatively, it could be the result of more complicated psychological phenomena, e.g. the social comparison of relative workrate (cf. Festinger, 1954). Again, we shall return to this issue in a more general discussion of our studies later in this chapter.

All of the findings thus far point towards fairness preferences in division and completion of work. One possibility is that the fairness we observe results from our very explicit framing of the allocation procedure using our game. It is possible that the prevalence of fairness is somehow resultant of our explicit framing of the planning phase of work; thus it would be interesting to explore whether similar results would emerge if an explicit framing was not used. We therefore designed a fourth and final study that replicated our previous experiments with a single change—our explicit framing of the allocation phase was removed altogether, and collaborators were simply instructed to organise the task as they saw fit. We examine how division of labour occurs without an explicit frame; whether searchers do indeed allocate work, and if they do, what allocations are made. We also use this next experiment in two additional ways. First, by introducing a new search topic, we examine whether the strategies we have observed also apply to other topics. Second, we examine whether the matching effect in search times occurs when our explicit framing is not used.

5.5.8 Study 4 Summary

This study was created in order to explore threat of rejection as an explanation of fairness observed in our division of labour ultimatum game. This was achieved by using a design in the spirit of the dictator game, where responders cannot reject the amount that has been allocated. We also introduced an information asymmetry manipulation where only proposers knew the amount to be divided. Through a study involving 40 participants, we found that:

- The modal allocation was once again an even split, occuring in 16 of 20 cases. Of the four notionally unfair offers, three were 4-6 and one was 3-7. Questionnaire responses indicate that a majority of participants invoked fairness when choosing how to divide the work.
- The information asymmetry had no noticeable effect on offers; dictators did not take the opportunity to increase their allocations in the knowledge that dictatees would not be able to evaluate them.
- When working on the Slime Mold search topic, participants once again used the Wait & Repair and Partition Web/Document Space strategies, but did not employ the Partition Semantic Space strategy.
- The search time matching correlation was replicated for a second time.
- A majority of participants felt that they needed to ‘keep pace’ with their partner’s work rate, particularly if the anonymous other worked quickly.

5.6 Thesis Study 5: Allocating in the Absence of the DLUG

5.6.1 Study Overview

In studies 2 to 4, we saw that fairness appears to be a guiding principle in the majority of our participants' interactions. One possibility is that the equality observed thus far might be an artifact that stems from explicit partitioning of the planning and acting phases of work. It would, therefore, be interesting to test whether similar behaviour would emerge when collaborators are not provoked into any type of formal articulation work—perhaps participants would not plan their activities so explicitly if the DLUG were removed. It would also be interesting to the allocation of work both quantitatively (*who does how much of what?*) and qualitatively (*who searches where for what?*), as well as whether we would observe the same evidence of fairness and equity as in prior experiments. The present study examines the veracity of our findings in the absence of our empirical frame. We retain our experimental task and invite participants to complete the work in whatever manner they deem fit. We also introduce a new information search topic to explore the use of coordination strategies as a function of the task at hand.

5.6.2 Design

This experiment was equivalent to studies 2, 3, and 4 in that we retained the scenario of dividing a task requiring 10 bibliographic references. This study again presented the task as forming a reading list for a night class. We retained the topics of pop music and slime mold, as used in study 3, and added a third topic on the subject of 'international art crime' to probe for new types of coordination strategy. This topic was selected on the basis of work by Bailey *et al*. (2009) who found that, out of 20 topics taken from the TREC robust collection (Voorhees, 2006), international art crime was perceived to be a difficult search topic. We expected that including this new topic might allow us to observe new strategies for coordinating work. Including the slime mold and pop music topics also allows for some superficial comparisons between the results of this experiment and studies 3 and 4.

5.6.3 Materials

All equipment and software were identical to those used in studies 2 to 4. The procedure and experimental protocols were similar to previous studies, save the necessary modifications required to remove all mention of the DLUG. (Materials can be found in Appendix F.) The instructions presented the experimental task in the same manner as before (*You have been asked to form a reading list for students of a night class...*). All questions pertaining to the DLUG were removed from the post experiment questionnaire.

We also added several questions to explore some of the issues that arose during earlier studies. Participants were again asked to indicate whether they felt influenced by their partners, and whether their decision to include sources was influenced by their partner. As in our earlier experiments, Likert scales were used to gauge topic familiarity, topic interest, task difficulty, and task interest. We also included the question used in study 4 regarding participants' perceptions about the need to speed up or slow down.

5.6.4 Participants

Thirty-six (13 male and 23 female) students, research, and administrative staff from 14 different University departments took part in this study. Participants' ages ranged from 18 – 46 (M = 26.7, SD = 7.1986). Demographic variables were not controlled for. This study did not involve any kind of role assignment (unlike previous studies, participants were not explicitly characterised as proposers and responders). Assignment of participants to pair was a matter of timetabling convenience.

There were six pairs in each condition; in the pop music condition, five pairs were mixed and one was female-female. In the slime mold condition, four pairs were mixed, with the others female-female. In the art crime condition, two pairs were female-female, one was male-male, and the rest were mixed gender pairings. None of the participants were involved in our previous studies. We again evaluated computing skill and search experience using measures from Morris (2008). All participants reported conducting Web searchers daily. Fifteen participants reported expert computing skill, with 21 reporting intermediate. These findings again indicated that our participants were comfortable with Web search and the use of computers.

As in our earlier experiments, we gauged participants' perceptions about topic interest, difficulty and familiarity using 1–5 Likert scales. For pop music, participants considered themselves moderately familiar with the topic (M = 3, SD = 0.9), did not find it especially difficult (M = 2.5, SD = 1) and found it quite interesting (M = 3.9, SD = 0.9).

The topic of slime mold was not familiar to most of our participants (M = 1.5, SD = 1), was moderately difficult (M = 3.08, SD = 0.6), and was not particularly interesting (M = 2.5, SD = 1.1). These results are highly similar to those recorded in studies 3 and 4.

Finally, although the subject of international art crime was not familiar to most of our participants (M = 2.2, SD = 1.5) participants did not find it especially difficult (M = 2.5, SD = 1) and, on average, actually considered it the most interesting of the three topics (M = 4, SD = 0.95). This result stands in contrast to the findings of Bailey *et al*. (2009), where art crime was perceived to be the *least* compelling of 20 search topics.[24]

Participants' rated enjoyment of the task was similar between topics (pop music: M = 3.75, slime mold: M = 3.66, art crime: M = 3.9), matching the findings of study 3.

5.6.5 Procedure

The procedure used in this study was identical to that of studies 2 to 4, except that participants received only a single set of instructions and no DLUG or explicit allocation procedure was used.

[24] This falls into line with our earlier discussions concerning the difficulty of obtaining a universally interesting or dull search topic.

POP MUSIC				
Session ID	Agreement	Sources Found P1	Sources Found P2	Total
4.1	5-5 proposed	5	6	11
4.2	None	8	3	11
4.3	5-5 agreed	6	7	13
4.4	None	5	5	10
4.5	None	13	9	22
4.6	5-5 agreed	6	6	12
SLIME MOLD				
Session ID	Agreement	Sources Found P1	Sources Found P2	Total
4.7	5-5 agreed	5	5	10
4.8	None	5	5	10
4.9	None	6	6	12
4.10	None	5	5	10
4.11	None	4	10	14
4.12	None	10	6	16
ART CRIME				
Session ID	Agreement	Sources Found P1	Sources Found P2	Total
4.13	None	6	7	13
4.14	5-5 agreed	5	5	10
4.15	None	6	4	10
4.16	None	5	5	10
4.17	None	5	5	10
4.18	5-5 agreed	5	5	10

Table 5.7: *Total number of sources obtained by each participant, study 5. P1 = Participant 1, P2 = Participant 2.*

5.6.6 Results

5.6.6.1 Workload Allocations

Since we did not prompt any allocation behaviour in this experiment, it is useful to explore two issues: first, whether an explicit quantitative allocation was agreed in each pair, and second, whether such allocations remained as agreed during the process of work.

We first explored whether participants agreed an explicit allocation, and, if they did, what these allocations were. Table 5.7 shows cases of agreement alongside the actual number of sources obtained by each participant. It can be seen that while only a third of pairs agreed an allocation, all were made at 5-5. It is worth noting that not all of these allocations ended up at 5-5; in half of all cases, participants posted more sources than absolutely necessary.

5.6.6.2 Motivation for Dividing Work

Participants were asked what factors, if any, they considered when organising the work. Only two participants mentioned fairness, e.g., *"I decided 50/50 would be a good fair start."* (P1, session 6, pop music). Other participants' rationales were more task-focused, either mentioning the need to have diversity in the reading list; the need to organise the work while preventing redundancy (i.e. what coordination strategy to use); or the need to ensure reliability in the reading list.

5.6.6.3 Coordination Strategies

Table 5.8 shows the strategies employed by participants in this study, alongside those identified in studies 2, 3, and 4. The use of coordination strategies in this study was relatively infrequent in comparison with previous experiments. Twelve pairs (66%) did not use a coordination strategy and have been classified as using the *Wait & Repair* method in Table 5.8. Only one of these pairs incurred a duplication, which was later corrected during the process of work.

Regarding explicit strategies, one pair used the *Partition Semantic Space* strategy in the pop music condition. No other explicit strategies were used in this condition. In the slime mold condition, two pairs utilised the *Partition Web Space* strategy. No other explicit strategies were used. In the art crime condition, one pair used Partition Semantic Space and two pairs used Partition Web Space. The vignettes below illustrate pairs using the strategies; the first shows a pair agreeing to partition the semantic space, and the second shows a pair agreeing a 5-5 allocation while simultaneously negotiating a Web space division.

> [Pair 4.4, Pop Music]
> Participant Two: ok should we pick 2/3 artists and then find sources of info on them?
> Participant Two: so let's say the beatles, rolling stones and jimmy hendrix?
> Participant One: We could do see, what we find

> [Pair 4.7, Slime Mold]
> Participant One: Hi, how shall we go about getting the reading list?
> Participant Two: hi there... do you wanna do 10 and i do 10 then share, or do you prefer do 5 each?
> Participant One: Shall we each do 5, but so that they don't overlap, would you like me to do journals and you do webpages?
> Participant Two: can do, I don't mind

Additionally, in the art crime condition, we observed the following interaction after the pair had initially agreed Web space partitioning:

> [Pair 4.18, Art Crime]
> Participant Two: I am struggling here. What keywords are you using to search?
> Participant One: "art theft" and "art crime" so far
> Participant Two: Thanks I will try something similar

Although this is in truth a case of one participant helping another, it hints that *Partitioning by Keyword Space* is a new strategy which might be used for coordination.

		Coordination Strategy		
Thesis Study	*Topic*	*Wait & Repair*	*Partition Web/Document Space*	*Partition Semantic Space*
2	Design	8 (1)	2	0
3	Pop Music	4	1	5
	Slime Mold	8 (1)	2	0
4	Slime Mold	14 (2)	6	0
5	Pop Music	5	0	1
	Slime Mold	4	2	0
	Art Crime	3	2	1
	Total	46	15	7

Table 5.8: *Frequency count of agreed coordination strategies seen in studies 2 to 5, by search topic. Numbers in brackets signify an instance of breakdown caused by duplication of work.*

5.6.6.4 Task Completion Times

We again derived participants' search times from process logs. Unlike earlier experiments, participants in this study often obtained more items than were specific by the task description. As we are interested in examining whether participants stop around a similar time, we opted to analyse search logs on the basis of the time at which each participant each pasted their final reference of choice. (As opposed to their fifth reference, as was appropriate in the majority of cases in earlier experiments). Such a decision offers a more accurate representation of participants' end points, as determined by their decisions to terminate the pasting of references.

Figure 5.4 graphs the search time data and again shows a striking concordance between partners' task completion times. Again, a Pearson correlation was computed on log transformed data, and revealed a strong positive correlation, $r(16) = 0.878$, $p < 0.001$.

5.6.6.5 Search Time Effect: Questionnaire Responses

We again asked participants about the extent to which they felt influenced by the actions of their partner. Twenty-seven participants provided some information about the issues affecting their workrate and decisions to paste sources. Appraisal of responses suggested that they would fit the categories identified in study 4. Table 5.9 again shows the categorisation of statements by the same two independent coders, who this time incurred one disagreement which was resolved through discussion. Results are similar to study 4 in that many participants again felt compelled to adjust their work to 'match' the apparent efforts of their partner.

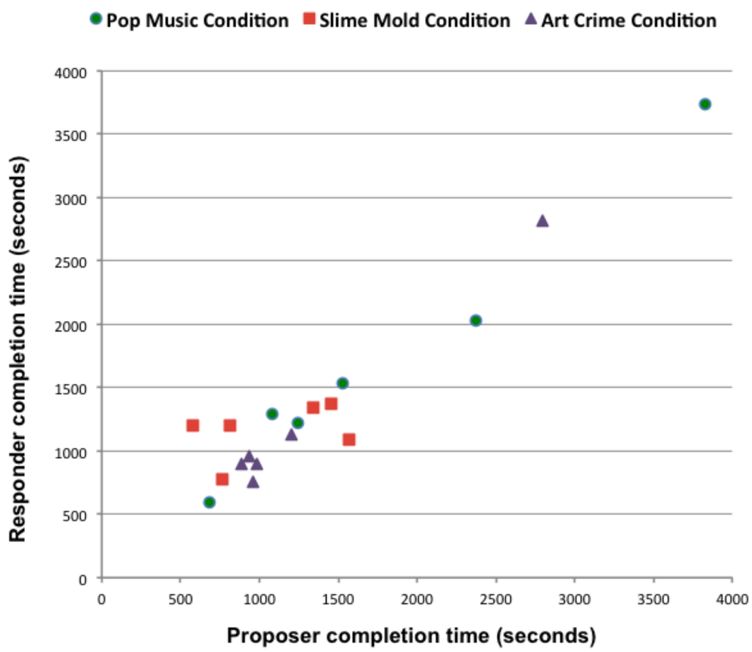

Figure 5.4: Correlation between task completion times of proposers and responders, study 5.

5.6.7 Discussion

This study sought to explore behaviour in our experimental task without using the DLUG to initially frame the allocation of work. We observed that relatively few pairs agreed, and subsequently followed through with, explicit allocations. Of those who did allocate, all were made at 5-5 splits. Participants' reported motivations for dividing the work in this way once again fall into line with fairness preferences. One interesting aspect of behaviour in this experiment was that 50% of pairs posted more sources than were absolutely necessary, i.e. they overshot the task requirements. This was not observed in our prior experiments. While this might initially seem like participants failed to properly internalise the task, we suggest that this behaviour has more to do with the way in which participants went about completing the task. That is, rather than define a stopping point for each person, participants collected references until they realised that the task requirements had been met. In some cases, the lack of clear division of labour resulted in unnecessary work. One perk of explicit allocations, as defined within the DLUG, is that it offers a clear objective for terminating individual activity. Indeed, other authors have argued that planned divisions of labour can serve as stopping rules (e.g. Bardram, 1997). The behaviour we saw in this experiment could actually be more representative of real-world collaborative search, where people often gather sources into shared repositories (or 'shortlists') and then discuss findings after an unspecified period of time (Kelly & Payne, 2014, see Chapter 6 of this thesis).

Turning to the enactment of work, the use of coordination strategies in this study was of similar frequency to that observed in prior experiments. Although it arose only once, we again find that *Partition*

Category	Description	Frequency	Example Statement(s)
Workrate Adjustment	Makes statement about adjusting their workrate based on that of their partner, i.e. indicating that they either sped up or slowed down.	11	*"They were working faster than me so I felt under pressure to go quicker. I would have liked to work slower to provide quality sources."* [P2, ID 4.8] *"Felt like a competition to find 10 links first!"* [P2, ID 4.11]
Coordination of Work	Makes statement about adjusting their search behaviour, i.e. altering their choice of pages visited or search engines used.	8	*"First three links by him/her were about americans, that made me look for artists in Europe/Asia."* [P1, ID 4.1]
Ensuring Quality	Makes statement about adjusting the way in which they judge the quality of found sources.	4	*"They seemed quite thorough so I tried to do the same."* [P1, ID 4.6]
Miscellaneous/Other	Makes statement with content that does not match other categories or statements.	4	*"Not really, he was going off track the majority of the time."* [P1, ID 4.5]

Table 5.9: Participants' responses to the question "Did you feel influenced by the activities of your partner during the study?", study 5.

Semantic Space is confined to the pop music condition, with those working on slime mold using *Partition Web Space* relatively more frequently. Of course, we must be modest about these results given the low number of observations, but the reoccurence of the strategies does at least strengthen our overall understanding of coordination during collaborative information seeking. We also saw that these strategies were used during research on 'International Art Crime', where a hint of another strategy, *Partitioning by Keyword Space*, was witnessed. It is likely that, with further work, other such methods of coordination would be identified.

We again observed a strong positive correlation in paired task completion times. This falls into line with the results of studies 2, 3, and 4. Both aspects of the effect were replicated: paired participants tended to finish at similar times, and there was considerable variance between some pairs (see Figure 5.4). We should note that the matching of times within a pair may initially appear less interesting because participants did not play the DLUG and were not incentivised with the opportunity to leave the study early. However, the replication of the effect here is important for two reasons. First, the fact that the effect persists without the DLUG suggests that it is not somehow tied to our explicit framing of the planning phase. Second, this study also shows that the matching effect need not be tied to a 5-5 allocation. If it were, overall times in the present study (where 50% of pairs involved participants each obtaining a different number of resources, e.g. 6 and 7) would not match up at all. The fact that participants complete different quantities of work, but still finish up at similar times, speaks to the overall consistency of the effect in the face of variable amounts of completed work.

Since the experiments reported within this chapter were essentially variations on a theme, and touch on the same issues multiple times, the following section of this thesis reflects more broadly on the present studies. We shall discuss salient issues that have come to the fore when examining bargaining over division of work. We shall also offer potential explanations and alternative interpretations of our various results, alongside some reflection upon the issues raised by our method.

5.6.8 Study 5 Summary

This study removed the DLUG and had no explicit delineation of planning and acting in division of work. The study was intended to explore the persistence of our results when our empirical frame, the DLUG, was removed from the division of labour task. We also introduced a new search topic, International Art Crime, to examine whether our observed coordination strategies map to other search topics. Through a study involving 36 participants, we found that:

- Most participants did not agree a quantitative split of the work. This resulted in many pairs doing more work than was absolutely necessary, pointing to the utility of explicit allocations as stopping rules.
- When participants did allocate (33% made an allocation), such allocations were made at 5-5.
- Participants once again used the Wait & Repair and Partition Web/Document Space strategies, and also employed the Partition Semantic Space strategy when working on the pop music and art crime topics. We observed a hint of a new strategy, Partitioning by Keyword Space.
- The correlation in search times was replicated for a third time. This indicates that the effect is not somehow tied to our explicit delineation of planning and acting via the DLUG.

5.7 General Discussion of Studies 2 to 5

The following subsections offer broader discussion of, and reflection upon, the present findings. In discussing our results we will consider in turn the DLUG allocations, the observed coordination strategies, and the matching effect in completion times. We will sketch possible explanations, provide further insights, and consider alternative interpretations as appropriate.

5.7.1 DLUG Allocations

A major aim of the experiments reported in this chapter was to explore bargaining over workload in our novel DLUG paradigm. In our first two experiments, we found that the modal outcome was a fair split (86% of cases), occurring when the search task was unknown, known to be interesting, or uninteresting to participants. We also found that a majority of proposers (73%) used fairness as a complete explanation of their offer. In our third experiment, we drew on the dictator game to explore whether proposers would become more selfish when responders had no recourse and limited information about the task. Even in this case, 80% of splits remained fair. Our last experiment explored allocations in the absence of an explicit frame. While quantitative allocations were somewhat infrequent, those who did explicitly partition their task typically did so at the point of fairness. Taken together, our four experiments provide further evidence for a fairness norm in the division of labour. Although it has not been our aim to compare the DLUG with the UG, one might actually argue that the frequence of equity is even stronger than that observed during negotiation over pecuniary payoffs. While our findings should be interpreted with caution given the number of participants involved, the results are at least stable and consistent.

The work we carried out during these studies was highly exploratory and, as such, was always a risk in terms of the results we might obtain. When conducting our studies, we were initially surprised (and sometimes disappointed) by the fact that participants did not attempt to minimise their workloads and instead repeatedly made fair allocations. However, considering our results in conjuction with relevant literature would suggest that the findings are, perhaps, not so surprising. In Chapter 2 we recognised that fairness is fundamentally about conforming to agreed norms, which may or may not be implicit, while ensuring that rewards are reflective of contributions. This was reiterated in Chapter 4, where, during our review of ultimatum games, we noted how monetary allocations decrease when proposers perceive themselves as entitled to more of the resource (as when claim is legitimised by an earnings task). Such findings are usually interpreted as ‘increased self-interest’ by economists, but are perhaps better regarded as demonstrating compliance with an expectation for fairness, at least in the sense that when only the proposer has exerted effort to obtain the reward, it is arguably fairer for them to be entitled to a greater share of the surplus.

In the present experiments, we tried to provoke a range of allocations by manipulating the attractiveness of the search task and by allowing proposers the opportunity to allocate without recourse. Regarding the former, participants with the unattractive task did acknowledge that they knew less about the task, perceived it as harder, and found it more difficult to obtain sources. Yet these factors did not impact allocation behaviour. One potential explanation here is that the task unattractiveness was not a strong enough incentive to encourage inequity. Yet perhaps a better explanation is that the incentive of a 'reduced workload', and hence opportunity to leave the experiment early, is not enough to warrant breaking fairness norms. Such an explanation might also be applied to the results of our dictator experiment as well—perhaps the disutility associated with being 'the bad guy' outweighs the benefit associated with delegating all of the work. These intuitions rest on the fact that participants do actually see the opportunity to leave the study early as an incentive. We must remember that participants took time out of their day to come to the experiment, and thus might actually be curious about the task and willing to try their hand at it. Delegating all of the work would deny such an opportunity, and might feel odd given that the participants were likely expecting to have to do 'something' to receive their reward. Issues such as these, as well as the highly exploratory nature of this work, make it important to consider *why* participants might be so inclined to make fair splits in the DLUG. We now consider some possible explanations, as well as alternative interpretations.

5.7.1.1 Equality as a Decision Making Heuristic

In discussing fairness in Chapter 2 of this thesis, we saw that people often make equitable allocations of goods as an efficient means of negotiating decisions in social contexts (Harris & Joyce, 1980; Allison & Messick, 1990). This 'general equality algorithm' (Hertel *et al*., 2003) serves as a fast and frugal heuristic for distribution of resources, especially when the decision involves an element of risk or requires extensive deliberation (cf. Collett, 1977). While minimisation of risk could play a role in invoking fairness (the equitable allocation is hypothesised to be an offer that mitigates the primary risk of the decision scenario, i.e. the threat of rejection), study 4 of this thesis demonstrated that risk minimisation alone cannot account for fairness in the DLUG. This is because participants continued to make equitable allocations in a dictator version of the DLUG where allocations could not be rejected. If risk minimisation had been the motivating factor in driving allocations, we would have expected to see many more unfair allocations of workload than were recorded.

This finding does not, however, completely rule out the use of the equality heuristic as an explanation. Participants might simply have been relying on equality for other reasons. In the context of our DLUG, an equitable allocation is straightforward in terms of cognitive processing, provides a reasonable reduction of workload, and conforms to egalitarian norms—equality is the default expectation in social exchange if no other mitigating factors are present. It may be that, rather than looking for the optimum reduction of workload, participants were instead looking for a 'good enough' outcome and considered fairness to be the 'just' option in light of its socially satisfying character and relative ease to obtain. Alternatively, it may be the case that participants actually want to try their hand at the work and thus have no strong desire to behave selfishly—this taps the broader issue of whether reduction in work is a sufficiently strong incentive to drive selfish behavior. What is interesting is that our participants might have been relying on fairness as a convenient heuristic even if it did not reflect their initial preference. Some individuals in study

4 stated that they had wanted to make an unfair allocation, but chose not to, instead opting for an even split. Exactly *why* participants chose this approach is an open question. We can offer an initial exploration by considering issues raised in the economic literature regarding incentives and moral costs in decision making. Before engaging such considerations, however, we first discuss two other plausible explanations.

5.7.1.2 The Role of Communication: Anticipating Collaboration

A better explanation of the results might be that participants were aware of the fact that they would need to work with their anonymous counterpart after making the allocation. In a typical UG experiment, participants come into the laboratory and are paired with an anonymous person in a separate room. No verbal communication occurs between the two, with the only information exchanged (at least in a baseline treatment) being the proposer's allocation and the responder's decision. This information is usually transmitted on paper or via computer terminal (Croson, 2005).

The procedure described above is one of the ways in which our DLUG experiments differ sharply to the classic UG, at least in procedural terms. Recall that participants in our experiments make their allocations via text chat, and were later free to communicate with their partners during the experimental task. While we made a reasonable effort to incorporate the methods of experimental models into our own studies, it was necessary for our participants to communicate in order for their collaboration to be at least something more than entirely artificial. One issue here is that communication is known to have a strong, positive impact on cooperation in ultimatum and dictator games (Frey & Bohnet, 1997; Bohnet & Frey, 1999; Croson *et al.*, 2003) and this is also true of other economic games (Sally, 1995). Since participants in our studies knew that they would be working with their anonymous partner, they would, presumably, have recognised that they might need to communicate with them during the experimental task. One might therefore speculate that the *anticipation* of collaboration explains the high number of fair offers in our DLUG.

Why might the very anticipation of collaboration result in a fair distribution of work? One explanation is that, by making a fair allocation of work, one is conforming to expected norms and is offering a very clear signal of willingness to cooperate. This aligns with the concept of *collaboration readiness*, which refers to the preparedness of an individual or organisation for collaboration, and includes both hard and soft factors (Rosas & Camarinha-Matos, 2009). Examples of hard factors include technological fit or matching competencies, whereas soft factors incorporate behavioural norms, values, traits, and trust (Rosas & Camarinha-Matos, 2009). Fairness, then, might be regarded as a soft aspect of collaboration readiness; by demonstrating that one is willing to behave fairly, one is demonstrating good character and providing a productive foundation for collaboration. Existing accounts of collaboration and the collaborative process (e.g. Roschelle & Teasley, 1995; Pinelle *et al.*, 2003) do not really consider fairness as a critical aspect of work. The present research, however, implies that fairness preferences may play a mediating role for collaboration from beginning to end.

The idea of 'anticipating collaboration' as an explanation of fairness could be readily explored using our DLUG. Such an experiment might contain two cells: one that involves subsequent collaboration after completing a DLUG, and one that does not. (The latter example would provide a perspective on the allocation of workload in terms of notional work units, even purer than that of our first experiment.) Such a study might also include a cell where participants engage in a standard UG over money, thus making the study a stepwise-esque comparison between a DLUG involving collaboration, a very rarefied DLUG, and the UG itself. Capturing a significantly different offer distribution with the non-collaborative DLUG would then give credence to the explanation offered above.

5.7.1.3 Collaborative versus Competitive Framing

An issue that is somewhat related to the preceding discussion concerns the way in which our DLUG is framed to participants. Ultimatum games are typically presented to participants as ‘decision making experiments’. Instructions and procedures are designed to abstract away from everyday scenarios and not activate “unconscious, preprogrammed rules of social exchange behavior” (Hoffman *et al*., 1996a, p. 659). In our studies, the framing is somewhat less neutral in that participants are aware that they are about to engage in a collaborative activity and will achieve some benefit for doing so (i.e., they can reduce their workload, leverage the benefits of having a partner for an unfamiliar task, and so on). Perhaps our task is perceived as an explicit opportunity to collaborate, rather than compete, with another player, and this framing gives rise to goals that do not correspond with the minimisation of work. Prior work in economics has shown that the way in which a game is framed can impact allocation behaviour: Larrick & Blount (1997) demonstrated the effects of framing the UG using the language of ‘claiming’ versus ‘dividing’, with the former producing more prosocial responses. Liberman *et al*. (2004) found that labelling a structurally equivalent prisoners’ dilemma as either a Wall Street Game or a Kibbutz Game[25] led to a decrease in cooperation when the competitive norm-inducing Wall Street frame was used. Since cues of social norms can enforce social norm compliance, the fact that ultimatums are not usually framed in terms of collaboration may, therefore, explain the prevalence of fair offers in our studies.

Our experiments were effectively labelled as a collaboration problem, and this means that participants may have been more cooperative than they would if the work was divided but no collaboration happened. While our decision to frame the experiment in the language of collaboration may explain the frequency of fair offers, we regard this as a positive aspect of our design that would only be of concern to the most hardheaded of economists. Our aim with this work has been to study collaborative division of labour; therefore, one might argue that evocation of relevant norms would actually be a *desirable* characteristic for our studies. Moreover, when introducing our model in Chapter 4, we introduced the workload minimisation approach as a baseline for DLUG play, intended primarily as a way of positing what players *could*, rather than *should*, do in the DLUG. We would argue that our results using a ‘collaborative’ frame are actually more valid in terms of investigating division of labour than an experiment that artificially imposed a

[25] Liberman *et al.*’s experiment was run in Israel: a Kibbutz is an Israeli collective community where the principle of equality is held in extremely high regard: goods are shared between all members of the Kibbutz, gifts are held in a common treasury, and all purchases are made collectively. See Talmon (1972) for further reading.

competitive norm as a baseline. However, we do not disagree that the possibility of evoking different behavioural norms is intriguing: normative framing has been shown to impact fairness preferences in both economics and psychology. For example, fairness judgements are more prosocial when people believe they are interacting with ingroup versus outgroup members (Hertel *et al.*, 2002, Study 1), and when ingroup norms are cooperative rather than competitive (Hertel *et al.*, 2002, Study 2). In line with our earlier discussion concerning the impact of communication on allocation, we would offer the issue of normative framing as an area for future work.

5.7.1.4 Moral Costs

A third consideration is the extent to which participants' decisions might have been impacted by factors that are known to impact decision making more generally. The issue of morality is especially relevant here: Levitt & List (2007) argue that, in any allocation, the utility derived from making the selfish choice has to outweigh the moral cost associated with the action in question. We hinted earlier that the moral cost of the selfish allocation in our experiments might have outweighed the corresponding gain in utility associated with doing less work. Our participants were given a very stark choice between completing an amount of work (of their own choosing) or, in a nutshell, freeriding in order to receive their reward. No participants preferred the latter option. Several stated that they 'felt bad' about making the selfish allocation, and instead opted for an even split, presumably to comply with social norms. Thus the potential disutility associated with selfishness appears to have been more salient than the payoff associated with a reduced workload, in turn explaining participants' gravitation towards fairness. This again raises the issue of work as a commodity in bargaining—we now address this issue in more detail.

5.7.1.5 On the Utility of Allocating Work

Our application of the ultimatum game to study division of labour has raised a number of issues about the act of deciding how to allocate a given task, and necessarily provokes broader questions about non-pecuniary bargaining. In particular, we can begin to think not only about what it is that participants are trying to maximise or minimise in our studies, but also our incentive structure and the relative utility of contributing work items to a joint project. These issues are interrelated to an extent, and discussion of each may assist in terms of explaining the frequency of equitable offers observed in our game.

Let us first consider the issue of what it is that participants are trying to achieve when allocating work in our DLUG paradigm. This question is intimately connected to the way in which participants value, and hence derive utility from, the commodity over which they bargain. As we saw in Chapter 4, a majority of subjects in economic experiments, like those that employ the UG, are incentivised usual actual financial rewards. The game-theoretic analysis of such experiments relies on the non-satiation postulate (Smith, 1976). This postulate assumes that, if given a costless choice between two alternatives, where the first yields more utility than the second, the utility-maximising option should always be preferred (Smith, 1976). In other words, if more money means more utility, then more money should always be preferable to less.

The incentive structure in a standard UG is, therefore, quite clear: proposers are assumed to be motivated to retain the surplus, and, for responders, the negative financial consequences associated with refusing the proposer's offer provide a clear disincentive for rejection.

If our experiments have shown one thing, it is that applying the utility-maximisation framework to division of labour is not straightforward. Unlike traditional economic experiments, our currency of interest is work, and one implicit assumption in our model is that less work is preferable to more. While we introduced this primarily as a simplifying assumption in Section 4.2, the analysis was intuitively sensible because if one is given a costless choice between receiving a reward for completing either more or less work, one might opt for the latter option as it maximises reward in relation to time. (Essentially, one is optimising 'wage'). However, participants in our studies did not behave in line with this reasoning. On the one hand, this makes our results qualitatively similar to the outcomes observed in typical UG studies, where empirical results fail to conform to the canonical economic model of behaviour. Yet on the other, our results beg the question of what exactly our participants were trying to achieve with their allocations. For example, when we attempted to vary the level of the tasks' desirability, as per our second experiment, several proposers actually kept *more* work for themselves. And later, in our third experiment, participants had the opportunity to delegate all of the work with no recourse, but most often chose to make a fair split. These results are counterintuitive given a work-minimisation model, and would seem to imply that subjects do not actually strive to minimise their workloads. The implication here is that participants clearly derive utility from other things than just minimizing work—defining exactly what those things might be presents a very clear and immediate opportunity for future work.

A second issue raised by this discussion concerns the nature of work itself. Our experiments essentially treat money and work as fungible, but the latter commodity is in many ways qualitatively different to money. Any currency lacks meaning if one does not appreciate, or cannot properly gauge, its value; the currency must, therefore, hold some intrinsic value to its bearer in order to be perceived as worth 'something'.[26] In a standard UG, the consequence of allocating more or less money is quite clear. In the DLUG, however, the exact value of the items in the pool to be distributed is uncertain: sources in a collaborative information seeking task are necessarily somewhat coarse with respect to actual effort, because, for example, one might expect some sources to be harder to find than others. Items in the pool are thus somewhat ambiguous in the sense that it may be difficult to predict how much time—and, by extension, effort—will be required to obtain each item. This in turn means that it may be difficult to assess the difference between alternative outcomes, making the allocation decisions we obtained potentially somewhat abstract—participants' allocations perhaps reflect an estimation of their and their partner's abilities to complete the task rather than their 'valuation' of the work. However, one might argue that the ambiguous nature of DLUG items is actually closer to the real world, where people are known to be bad at predicting task completion times (Buehler *et al.*, 1994). Allocation of any task, then, must involve

[26] See, for example, the money illusion effect, where consumers tend to rely on the nominal rather than absolute value (in terms of purchasing power) of money when evaluating transactions (Shafir *et al.*, 1997). When evaluating unfamiliar currencies, people spend less if the currency is a multiple of their home country, but more if it is a fraction, suggesting they find it hard to gauge the value of unfamiliar currency (e.g. Raghubir & Srivastava, 2002; Wertenbroch *et al.*, 2007).

some level of ambiguity, and attempting to fix or constrain this property of work could be problematic. If anything, our studies reveal the nature of work as a much more nebulous commodity than money. Since previous work has shown that ambiguity and risk can impact economic decision making (e.g. Heath & Tversky, 1991; Kahneman, 2003) future work might specify the time required by each subtask so as to investigate perceptions of notional versus absolute work units. Our decision not to specify an exact completion time for the experiment can be justified by the discovery of our search time effect, but expected time investment is another factor that could be controlled in future.

Our final consideration on this matter is that work is characterised by myriad parameters which may impact allocation choices. If one focuses on time spent working on a task, one might imagine preferring fewer work items. If one instead focuses on the quality of the final product, then delegating work to another requires trust in their competence. If one seeks to maximise efficiency, then one must pay attention to skill and efficacy, as was done by participants in study 3. And the fact that the work may not be aversive at all (some tasks are intrinsically motivating) means that the relative utility of contributing a work item to a joint project is unclear. These are all considerations which will affect division of labour, and might also be related to fairness in that there may be circumstances under which unfairness is legitimised by efficiency. Such considerations provide an exciting array of possibilities for putting our novel DLUG to use in other studies.

These considerations reveal that our exploration of the DLUG as yet only scratches the surface of its potential, and the numerous issues raised by this discussion should not necessarily be regarded as problematic. Rather, we view these avenues as potentially fruitful areas for further research. Our long-term hope in designing the DLUG is that it might allow insights into division of labour, planning, and negotiation in collaborative work more generally. And although we have manipulated how interesting a task is and found little effect on allocations, we remain optimistic that studying the impact of work characteristics on the DLUG, and comparing these with the well established phenomena in the classic UG, might allow the DLUG to throw light on the judged utility of work. We regard an understanding of the utility of work, and of components of projects, including the utility associated with work done by others rather than oneself, as a vital boundary of progress in the understanding of collaborative work, especially in the various trade-offs between delegation or distribution of labour and control. (Just as it is now recognised that considering utility is vital for understanding individual trade-offs in cognitive performance, e.g., in speed-accuracy tradeoffs (Howes *et al.*, 2009)). Similarly, the possibilities outlined in this discussion raise the possibility of exploring punishment strategies and framing effects in our DLUG. Chapter 8 of this thesis distils these various possibilities into a scheme for future work.

5.7.2 Coordination Strategies

We found that searchers used a variety of coordination strategies to organise their work. Uncovering these coordination strategies helps to justify our DLUG in terms of producing pragmatic outputs that can be of immediate use to HCI. Our foremost finding, beyond the strategies themselves, was that searchers may use different strategies to coordinate their work so as to avoid the problem of duplication. While studies 3 and 5 showed that search topic did not affect quantitative workload allocations to any great extent, we did observe that properties of the shared task influence the availability of coordination strategies. That is, participants working on pop music often chose to partition by semantic space, whereas this strategy was never observed in the conditions involving the slime mold and design psychology topics. However, the strategy was replicated in study 5 with the pop music and art crime topics. Perhaps the most valuable aspect of these results is that, as hinted in our introduction to this chapter, collaborative search systems might be designed to provide support for the division of labour policies we have identified. Our strategies are likely not the only available methods for coordinating work during collaborative search, and thus future research might continue to explore emergent coordination behaviour in the hope of identifying new opportunities for creating division of labour. More broadly, our strategies imply that designers might wish to consider different ways of partitioning work, as relevant for the task at hand. For example, in a system for collaborative travel planning, collaborators could partition work by separating potential destinations, transport versus accommodation, and so on.

Interestingly, we did not observe any participants postponing the management of duplication to a separate final stage, as suggested by the *brute force* strategy identified by Morris (2008), where collaborative searchers divided work and then amalgamated sources later, leading to redundancy. It is possible that this strategy is used more often in asynchronous collaborative work. We have obtained no evidence in our studies that the various coordination strategies are differentially effective, but it seems plausible that some strategies will be better for some tasks than others. Future work might explore these issues by instructing teams to use various strategies, and by using more refined metrics to judge the quality of team performance, e.g. number of queries issued, or time on task in relation to strategy employed.

5.7.3 Search Time Effect

Perhaps the most striking result from the present experiments was the close correspondence of completion times of partners within a pair. This occurred regardless of any particular pair's completion time: some pairs finished the task quite quickly, whereas others took longer, but within-pair differences were highly similar across the board. Thus it seems that participants, for whatever reason, appeared to 'yoke' their efforts so as to end up with equity in the overall completion of the task itself.

Not only was this finding serendipitous, its occurrence was surprising given that participants were neither incentivised nor instructed to ensure similarity in their workrates. If anything, one might expect that such behaviour would be discouraged by many features of our experimental design: participants were anonymous and under circumstances of limited identifiability; were spatially separated and communicated only through text chat (meaning that awareness was fairly limited); and were free to leave the experiment after finishing their own allocation (implying that their work is in no way dependent on the partner's). Lastly, and perhaps most pertinently, there seems to be no real *reason* to match completion times given that this pattern of behaviour confers no obvious economic or procedural advantage. Here we attempt to interpret the effect and consider its underlying cause, as well as how it came to pass in the present experiments.

In terms of understanding *why* our participants might have strived for equity in completion, several clues can be found in the literature on social loafing, which refers to the tendency to invest less effort when working in groups than when working alone (Latane, Williams, & Harkins, 1979; Kravitz & Martin, 1986; Kraut, 2003). Slacking behaviour of this nature has been demonstrated in a variety of tasks requiring either physical or cognitive effort (including shouting, rope-pulling, and brainstorming; see Jackson & Harkins, 1985, page 1199). Social loafing is typically regarded as a negative phenomenon since it most likely leads to a decrease in overall productivity for groups. At least two variables are known to mediate social loafing. First, the ability of participants to identify each individual's contribution appears to be key: participants are seen to loaf if they believe others will not bear witness to their efforts (Latane *et al.*, 1979) but the loafing effect disappears if mutual identifiability is held constant (Williams *et al.*, 1981). Second, researchers have shown that social loafing occurs when individuals *expect* their coworkers to loaf—such expectations cause people to subsequently lower their own efforts so as to "establish an equitable division of labour" (Jackson & Harkins, 1985, see Abstract). While both of these explanations aim to account for observed shortfalls in collective effort, research has also shown that efforts can be *increased* according to expectations. For example, Jackson & Harkins (1985) found that when participants worked with a confederate and expected her to invest high effort, participants followed suit, but when the confederate was expected to exhibit low effort, participants also invested low effort themselves. Since these results suggest that effort can be increased *and* decreased according to knowledge about shared exertions, social loafing can actually be regarded as the lower end of a more generalised effort matching phenomenon.

In their attempts to account for these effort matching effects, Latane *et al.* (1979) suggest that participants adjust their effort levels so as to ensure that work is completed in line with fairness norms. Elaborating on this explanation, Kerr (1983) suggested that if one's partner exhibits (or is expected to exhibit) loafing behaviour, one might lower his or her own effort so as to avoid looking like the 'sucker' (as opposed to raising effort so as to compensate for the shortfall in productivity). Such arguments are clearly of relevance to the present thesis—the effect we tapped here might simply be this 'effort matching effect' at work, in turn providing further evidence of a general desire for fairness in the process of completing work.

An alternative explanation for social loafing, and the effort matching effect more generally, is that participants engage in social comparison so as to determine how much time to spend on a given task. Social comparison theory (Festinger, 1954) posits that people frequently compare themselves to others in order to evaluate their own opinions, abilities, and actions. As articulated by Jackson & Harkins (1985), this means that people may "match their level of effort not to the partner's level, but to some normative standard that is communicated through the partner; that is, the partner's expressed intention to work hard or hardly work gives the participants information about how people, in general, would respond. Thus participants match effort not because of a sense of fairness or equity, but simply because they are conforming to some standard" (Jackson & Harkins, 1985, p. 1206). While we recognise that this explanation is not immediately applicable to our work (if only because neither one of our participants is an experimental confederate masquerading as a coworker), social comparison effects might still be responsible for our own results in a similar, but subtly different, fashion. This is described as follows.

Research has shown that people are especially prone to engage in social comparison during circumstances where behavioural requirements are ambiguous or unclear (Karau & Williams, 1993; Gibbons & Buunk, 1999; Mussweiler, 2003). Our experiments did not specify exact requirements for time or quality of work, meaning that participants were left to determine how to satisfy these vague criteria for themselves. Perhaps in attempting to resolve this ambiguity, participants used cues about one another's workrate to arrive at some emergent normative standard, in turn leading to eventual similarity in completion time.[27] Feldman (1984) argued that norms are often determined by initial patterns of behaviour when a group comes together; tentative, early actions mark the boundaries of acceptable behaviour and lead to expectations about performance. Since norms can spontaneously arise when behaviours that satisfy one group member are imitated by others (Opp, 1982), our participants might have finished up at similar times simply by pacing themselves at some mutually satisfying rate. That is, rather than striving to meet fairness norms, the norm was simply to 'do as my partner is doing'. Since none of our participants engaged in any explicit discussion about how long to spend on the task, any such matching effect must have emerged implicitly rather than via explicit negotiation—a feature that makes our observed correlations all the more fascinating. Furthermore, it is difficult to imagine that every pair would arrive at the same normative standard for effort investment, meaning that we would naturally expect some variation between pairs if this explanation were true. And, as it turns out, our work offers strong support for this intuition by the fact that completion times in our studies were often dissimilar between many of our pairs.

In terms of how this process might have been bootstrapped in our experiments, insight can be obtained by considering the way in which participants were able to interpret one another's efforts during our experiments. Recall that the only indication of workrate available to each participant was the pasting of references by their partner into the shared chat. This shared resource was in fact the sole awareness mechanism provided to our participants. The qualitative responses collected in studies 4 and 5 indicate that participants sometimes felt a need to increase their workrate due to the perception that they were 'lagging behind', particularly when the anonymous collaborator seemed to be working quickly. Others testified that the work process sometimes felt like a 'competition' to see who could finish first. It would, therefore, seem that if one participant arbitrarily decides to work quickly, the second feels the need to speed up, hence leading to a matching effect. This would certainly help to explain why some pairs finished the task in a much shorter timeframe than others. In some cases, it is possible that the matching effect might be the result of social pressure—the perception that one is 'lagging behind' very likely leads to an upturn in completion time (perhaps facilitated by social comparison as well, at least to some extent).

[27] Levine & Moreland (1998) define norms as "shared expectations about how all group members ought to behave" (Levine & Moreland, 1998, p. 427). Norms are a characteristic of group structure that can guide how a group goes about its business, and can be determined formally or informally. Norms are sometimes inferred from the behaviour of other group members, with regularities interpreted as conformity to group norms and irregularities interpreted as deviance (conflict) from those norms (Levine & Moreland, 1998).

However, this interpretation does not provide insight as to why matching occurs when no such pressure is felt, as is presumably the case in longer trials. Perhaps in this latter case behaviour is based on judgements of partner competence, as evidenced by statements like *"they seemed quite thorough so I tried to do the same"* [Participant 12 in study 5]. Whatever the cause, it appears that the general phenomenon is one of matching to some perceived standard—the question for future work is how such standards are derived. One issue that might play a role is intrinsic motivation regarding the task. That is, if an estimation of the other person's effort is the sole driver on one's own investment, then the effect is likely to be normative and a correlation would emerge. However, if the effect is driven by issues such as independent interest in the search topic, then perhaps we would not see a yoking effect. Providing proper answers to these questions clearly requires further experimentation.

Although we have expressed surprise at the yoking of task completion times, it is possible that the effect we observed is merely a reduced form of how groups naturally arrive at pacing of their work. The literature on small groups uses the term *entrainment* to describe the way in which groups repeatedly adopt the same workrate across temporally distinct tasks (Kelly & McGrath, 1985). If a group is given 10 minutes to complete a task, and then later is asked to complete a similar task with a 20-minute deadline, group members tend to work at the same rate as occurred beneath the shorter deadline (Kelly *et al*., 1990). In settings with time abundance, groups have to set a workrate for themselves. Workrate likely emerges as a normative standard, based either on prior example or on some initial tendency that is reached arbitrarily (perhaps by one member deciding to invest high effort and others following suit, as suggested above). What would be interesting is to see whether pairs who invested a high or low amount of effort in our paradigm also did the same when invited back for a second task, or whether the behaviour of persons who invested little effort could be shifted by pairing them with a thorough searcher. Any new emergent effects, alongside participants' testimonials concerning decisions about workrate, could then throw greater light on some of the questions we have posed during this discussion.

Lastly, it is worth returning to the issue of how participants determined an appropriate level of effort in order to understand *how* the phenomenon might play out pragmatically. In the present experiments, we required each participant to set his or her own threshold for what constituted a suitable contribution for the shared bibliography. As far as our experimental protocol is concerned, any source is acceptable—pay is not based on performance, and participants are not disqualified from payment for producing poor or irrelevant sources. Perhaps the matching effect can be explained in terms of participants adjusting the threshold of acceptability in accordance with the workrate exhibited by their partner. In other words, a participant faced by a partner who appears to be working quickly might respond by adjusting their own acceptability threshold downwards, allowing them to speed up the pasting of their contributions and keep pace with their partner. Likewise, if work begins at a leisurely pace, perhaps the threshold for analysis remains constant or is raised as contributions are added. We should note that although we have measured and described the matching effect in terms of completion times, it is entirely possible that it emerges from participants attempting to match some other aspect of their efforts, such as the resulting quality of found sources, and we note that several trials did involve negotiations about what might constitute an acceptable level of quality for the reading list. However, this cannot explain why the matching effect occurs in trials where no communication occurs whatsoever, as was the case in several of our experiments. Perhaps some participants were content to match time, whereas in others quality was also a criterion. As we did not collect data that would allow us to resolve these considerations, further research is required to delineate these issues.

While the present discussion has attempted to account for the effort matching phenomenon observed in our experiments, it is quite clear that further work is required to disentangle the exact mechanisms underlying the effect. Irrespective of its cause, it appears safe to say that people seem to strive for some degree of equity when completing shared tasks. The following subsections consider the theoretical and practical implications of the search time results recorded in the present experiments.

5.7.3.1 Implications for Theory

Prior work on effort matching effects typically employs experimental confederates in order to bring about changes in participants' efforts. For instance, the study by Jackson & Harkins (1985) used an assistant who, in a high effort condition, said she had "tried very hard on the practice trial and, because she thought the research was interesting, was going to try hard throughout the experiment" (ibid, p. 1202). In the low effort condition, she stated that she "had not tried hard in the practice [trial] and, because she thought the research was boring, wasn't going to try hard for the rest of the experiment" (ibid, p. 1202). Participants were later seen to shift their efforts so as to match the expressed intention of the confederate; the implication is that expectations about one's partner bring about changes in one's own behaviour, as mentioned above in subsection 5.7.3.

Our findings are made distinct from prior studies by the fact that we do not use experimental confederates but still witness matching behaviour from our participants. As discussed earlier, we did not specify exact requirements for effort, meaning that participants in our experiments are responsible for setting their own effort threshold. The present studies might be regarded as evidence that effort matching can arise naturally and in a non-contrived fashion. Furthermore, the present studies suggest that effort matching can arise *without* explicit knowledge of a partner's intentions—we remarked earlier that none of our participants discussed how long to spend on the task, and, in many cases, times are highly similar without any conversation occurring at all. Perhaps effort matching is a pervasive phenomenon in collaborative work groups—further work should explore whether a similar effect can be obtained in non-laboratory contexts.

5.7.3.2 Implications for Collaborative Systems

An exciting implication of the search times finding concerns the role that technology may have in mediating the effect. In particular, one might imagine that the effect is moderated by the level of awareness provided by the collaborative software. In the present experiments, awareness was maintained via the pasting of references into the chat; participants were not actually able to see what was happening on their counterpart's screen. Yet even this relatively meagre level of awareness was enough to facilitate effort matching. A more design-oriented implication, then, is that the level of awareness participants have in a collaborative system may affect not only their ability to collaborate, but might also influence their ability to make fairness judgements and to pace their overall workrate. Perhaps varying the awareness information provided to participants might disrupt, or even enhance, the matching effect. One would expect that without the ability to have at least some awareness of joint action, the synchronisation effect would disappear completely—it would be impossible to match effort without the ability to make ongoing judgements about a partner's workrate. Similarly, if the 'paste as you go' approach favoured in our experiments were removed in favour of delayed pasting, i.e. sources are only entered after a searcher has found enough to satisfy his or her contribution, then the matching effect is likely to disappear. Both of these points are more logical observations rather than empirical questions, but it would be intriguing to explore how the synchronisation effect might be perturbed by more or less granular levels of awareness in whatever technology participants happen to be using at the time.

5.7.4 Task Completion Time and Quality of Work

Although our experiments saw variable levels of investment from our pairs into the shared task, one question not explored in our analyses was whether time on task equates to a better quality of work product. That is, we do not know whether pairs who chose to work for longer periods of time acquired sources that were more relevant to their assigned search topic than those who worked quickly. Analysis of participants' work quality might further our understanding of why some spent longer on the task than others—although we did not specify a precise expectation for effort, perhaps some participants were evaluating sources more thoroughly in an attempt to provide high-quality contributions to their bibliography. If this were the case then we might expect those with longer completion times to have finished up with better, more relevant sources than those who worked quickly.

One might intuitively expect that spending longer on a task leads to a better quality of output, and evidence from the literature on small groups supports this notion. A classic finding is that time constraints lead to a performance tradeoff, where time-pressured groups tend to work at a higher rate of output than those who are not time pressured. However, this increased productivity comes at the expense of both quality and quantity of work products. For example, Kelly & McGrath (1985) found that groups working on writing tasks were more productive (in terms of words-per-person-per-minute) with a 10 minute time constraint, but the work completed by these groups was shorter in length and of lower quality than that produced by groups working beneath a 20 minute deadline. Other authors have obtained similar results when studying different group tasks (e.g. Karau & Kelly, 1992; Kelly & Karau, 1999; Harrison *et al.*, 2003; Kelly & Loving, 2004).

To examine the relationship between task completion time and work quality in our studies, we performed exploratory analyses using the results of Thesis Study 2 (our first DLUG experiment) . Recall that the search topic from the experiment was: *"To what extent can design be considered a psychological* *process?"* and pairs had to deliver a total of 10 sources that could be used to answer this question. To determine the quality of each participant's found sources, we adopted the established procedure of using judges to evaluate the quality of work products (Kelly & McGrath, 1985; Harrison *et al.*, 2003). For the purposes of the analysis, we defined 'better quality' work as sources that were 'more relevant to the search topic'. Two independent coders with academic backgrounds in HCI were given complete lists of the references (69 total) collected by our participants. Coders were blind to the purposes of our study. We asked the coders to examine the content of each reference and provide quality ratings using a 1–5 Likert-type scale, where 1 = "Not at all relevant" and 5 = "Highly relevant". Inter-rater reliability was found to be .74, with disagreement denoted by a disparity of two or more data points between ratings, e.g. coder 1 rating = 2, coder 2 rating = 5. Because the two coders were unable to convene for discussion, a third independent rater was given all items of disagreement alongside a random subset of other references to ensure masking. We then used the evaluations of the third coder to resolve disagreements.

The quality score for each reference was used to acquire an overall quality rating for each participant's work. This was done by averaging the ratings of the five items collected by each person in the study (*Mean* quality rating = 3, SD = 0.7). We then correlated each individual's rating with their task completion time to explore whether those who worked for longer acquired sources that were of better quality. The analysis revealed a weak positive correlation which was not statistically significant, Pearson's product-moment r = .17, p = .473. This would suggest that spending longer on the information search task did not lead to better quality of found sources.

Although this result appears to go against prior evidence from the small group literature, there are two issues which need to be considered. The first is that previous studies use measures of quality that are quite different to ours. Several studies (Kelly & McGrath, 1985; Karau & Kelly, 1992; Harrison *et al.*, 2003) use scales developed by Hackman *et al*. (1967) in which quality is assessed according to various dimensions, including the *visual presentation* of a group's efforts. This scale was used as a measure of quality in the study by Kelly & McGrath (1985), in which the quality of presentation was higher when groups had an extra 10 minutes to complete their task. Visual presentation is a quality criterion that does not map well to our study: participants were simply pasting references into a shared chat window, and since our task description did not suggest a need for the bibliography to be aesthetically appealing, it would be difficult to distinguish our participants' efforts by visual presentation. A more appropriate criterion might be that of *adequacy*, which was used as a measure of quality in studies by Kelly & McGrath (1985) and Karau & Kelly (1992). The concept of adequacy seems to be closer to our concept of relevance, and what is especially interesting about this metric is that both of the studies mentioned previously found that groups who spent longer on a task did *not* produce more adequate solutions than groups who were time pressured.[28] It may thus be the case that extra investment beyond a certain threshold improves only certain aspects of a group's work and not others. For example, groups might use abundant time to improve quality

[28] In spite of our best efforts, we were not able to examine the original measure of adequacy used by Kelly & McGrath (1985) and Karau & Kelly (1992). This is because both studies point to the work of Hackman *et al*. (1967) as the source for the adequacy measure, but the cited paper makes no mention of the scale. It is difficult to know what to make of this beyond the suggestion that it may be an endemic mistake in the literature.

of presentation in favour of content. For the present thesis, the result perhaps emphasizes the fact that there was no real advantage to spending more time on the task, especially since effort was not a criterion for success in the experiment.

A second issue is that the mean quality of participants' findings tended to fall at the midpoint of the scale used by our coders to evaluate sources (M = 3, scale = 1–5). It is not the case that our participants were polarised in the sense of producing either extremely poor or high-quality work. Rather, the majority of bibliographies were something of a mixed bag, with each participant's contributions comprised of several sources that were good and several that were poor. This led the majority of ratings to be clustered around the midpoint of the scale. This could perhaps be caused by a learning effect where participants gradually developed a better understanding of the search topic and thus contributed better sources towards the end of their work. However, eyeballing the raw data does not suggest this to be the case: some participants pasted their least-relevant source last, whereas for others quality was variable throughout their work. Further insight on this matter could be gained by questioning participants about their decisions during the experiment, but this is unfortunately beyond our reach at this juncture.

One final point worth considering is that participants' work quality might have been impacted by their perceptions of what constituted an appropriate source for the bibliography. We asked participants to collect 'reliable sources'. Some of the gathered references were highly relevant to the topic (mostly those from scholarly journals) whereas others were only tenously relevant but could nevertheless be regarded as *reliable* due to the reputation of their source. For example, one participant contributed a BBC news story on colour psychology. This item was given a very low relevance rating by our coders but could arguably be considered as reliable because it originates from a trusted source. Our definition of quality might not be in line with that invoked by our participants—future studies that attempt to explore the relationship between quality and time should ensure that this confound is avoid and that expectations about effort are made clear to participants.

Overall, our analysis of Study 2 suggests that time on task did not lead to better quality contributions. This result is potentially interesting because it runs counter to the intuitive supposition and appears to fall into line with some prior evidence from the small group literature, but should perhaps be interpreted with caution due to the various issues outlined above. Since the studies in this thesis were not specifically designed with analyses of quality in mind, combined with the fact that appraisal of participants' sources by independent coders was both time and labour intensive, we elect not to pursue this line of inquiry any further.

5.7.5 Limitations And Further Work

The present experiments are limited in scope and thus myriad opportunities exist for further work. First, the experiments reported in this chapter are merely a drop in the ocean in terms of what could be explored using the DLUG. Chapter 4 explored thirteen factors that are known to impact behaviour in the classic UG—future work could continue to attempt replications of known effects, as attempted with the information asymmetry enacted in study 4 reported here. Alternatively, the DLUG could be used to explore factors relevant to the assignment of work; for example, status hierarchies, knowledge disparities, skills—the possibilities are quite extensive. In developing the present experiments, we were affected by factors including time constraints, limited staff support and constrained experimental budgets. It would be interesting to see what could be achieved if these issues were not present.

One immediate area for future work would be to explore how different variables can legitimise unfair, or at least inequitable, allocations. Our attempts in this regard were unsuccessful, but one manipulation that may have a noticeable impact is that of entitlement. Positions in the DLUG could be awarded based on prior knowledge, performance on an earnings task, or by allowing participants to place bids on roles (cf. Güth & Tietz, 1986). Such entitlements could be more or less distorted, e.g. with only one player completing the earnings stage. Participants might be more inclined to shift away from equity beneath such manipulations, and further work could then explore the impact of inequity on process results.

Although we attempted to mirror the procedure of economics experiments quite closely, there are a number of concerns related to our protocol that could have impacted our results. One is the issue of scrutiny, which can affect decision-making experiments in two ways: first, by the extent to which participants perceive their decisions as subject to inspection by referent others, i.e. the experimenter, and second, by lessening intersubject anonymity—participants may view their own role in the experiment with greater saliency if they do not see themselves as 'lost in the crowd'. Perhaps in our experiments, which involved a number of procedural differences to standard UG procedures, participants were acutely aware that their decisions might be scrutinised at a later time—the very presence of a facilitator might have caused participants to behave in the socially acceptable manner, especially if they thought the facilitator might bear witness to the allocation made using the chat window. Participants' decisions may therefore reflect what they thought was morally correct rather than their innermost preferences. One other procedural difference is that participants completed our experiment one at a time (in anonymous pairs), rather than in groups as is typical of economic experiments (Croson, 2005). While individual participation has been used in prior UG studies (e.g. Henrich *et al.*, 2004), it is possible that our participants felt highly individuated. Both of these issues warrant further study.

A second procedural issue concerns the impact of social norms in tandem with the type of people recruited for our experiments. Our participants voluntarily self-select into our experiments, meaning that they might naturally be 'cooperative types' who are, on average, more prone to agreeable behaviours and are more likely to act in line with fairness norms (Levitt & List, 2007). This may correspond with a lack of lazy, free-rider types, who are unlikely to go out of their way to assist an experimenter for a relatively small reward. This in turn could have affected our results in that our samples are not properly representative. We would note that issues concerning self-selection are a broader concern for laboratory experiments and there is not much one can do about this beyond further research.

Lastly, one limitation of the present experiments is the use of synchronous collaboration in which participants performed their parts of the task at the same time, and this may have played a vital role in allowing them to match effort. It remains to be seen if asynchronous workers are similarly influenced by whatever cues they might receive about their partner's efforts. Further work might explore whether effort matching occurs in tasks beyond collaborative search. Clearly there is a need to explore whether fairness persists in more or less aversive tasks, and such studies can push on our contentions about the utility of work; for example, in a highly aversive task, participants might be much more tempted by the opportunity to delegate and avoid work. Alternatively, they might feel more risk averse or more guilty about palming off their perceived responsibilities. Further work can push on our ideas to develop utility functions relevant to CSCW that account for the various factors that impact division of labour during collaborative work activities.

5.8 Chapter Summary and Conclusion

In this chapter we explored our DLUG model in a series of empirical studies. Our model was designed to isolate the planning phase of work, and we used four experiments to explore how participants divided their labour and whether such divisions were taken in line with fairness norms. Our task of choice was collaborative information seeking, implemented as requiring collection of 10 bibliographic references. In summary, we found that:

- the modal workload allocation in all of our experiments was an equitable split of the work, and notionally unfair allocations did not stray very far from the point of equity. (6-4, and one case of 7-3). All of the allocations were accepted.
- participants used a variety of coordination strategies to enable division of labour and navigate the problem of redundancy. Examples include organising their work by dividing different search engines, or by semantic properties of the task at hand.
- there was an emergent matching effect in participants' task completion times, which may be evidence of a broader effort matching effect (cf. Jackson & Harkins, 1985) enabled through social comparison and the ambiguous effort requirements of our task. Such an effect can be interpreted as evidence for overall equity (and hence, fairness) in the work process (Latane *et al.*, 1979).

Taken together, the findings of the present experiments provide further evidence of preferences for fairness in division of collaborative work. Putting aside issues associated with our experimental model and the utility of work, we suggest that fairness may play an important signalling role in collaborative work—by proposing to divide work evenly, and then following through on that proposal, one is giving a clear indication of willingness to cooperate, in turn demonstrating collaboration readiness (Rosas & Camarinha-Matos, 2009). Fairness is also a very straightforward, cognitively inexpensive allocation (Allison & Messick, 1990) that creates a harmonious basis for collaboration—which is how our experimental task, with its suggestion of interdependency and a shared goal, was likely perceived by participants. Not only do these findings suggest theoretical merit for the DLUG, the coordination strategies offer direct input for collaborative search systems, in turn suggesting that the DLUG can be used to provide productive design implications as well.

The following chapter expands the scope of our concerns to the extra-lab world. We continue to study the task of collaborative information seeking, as well as the extent to which searchers divide their labour and desire fairness, but in the context of everyday search tasks. We report a field study where pairs of searchers used an existing collaborative search tool to complete everyday tasks of their own choosing (e.g. travel planning). The study is both theoretically and practically oriented in that we aim to understand real-world collaborative search while also expanding the contributions of this thesis for collaborative search tools.

CHAPTER 6

COLLABORATIVE SEARCH IN CONTEXT: A REAL-WORLD STUDY

6.1 Chapter Overview

Chapters 4 and 5 reported empirical work based primarily on results obtained from the laboratory, where fairness and division of labour were explored using collaborative search tasks. We also described how participants coordinated their work to prevent redundancy during search. Since those studies were, at least initially, geared towards exploring behaviour in the novel DLUG, this chapter takes a different approach by exploring division of labour and fairness outside of laboratory settings. Here we report a qualitative study that explores how pairs of searchers organise their work during everyday collaborative search tasks. The first motivation for this study is the need to consider how division of labour is played out in other situations, both in terms of how work is organised and the extent to which participants care about fairness.

A second motivation for this study is to enhance the pragmatic and design-oriented contributions of this thesis. We noted in chapter 5 that collaborative search is a situation in which division of labour is often desirable, and that the question of how best to implement division of labour is currently an active one within the research community (Foley & Smeaton, 2009; Morris, 2008; Kelly & Payne, 2013). In the present study, we deployed two existing collaborative search tools to pairs of searchers conducting everyday information seeking tasks. Our immediate aim was to understand whether the tools were useful in supporting actual collaborative search, and thus we recruited pre-established collaborators with information needs that were intrinsically collaborative—tasks were self-selected, and participants used their assigned system for as long as they wished, creating high external validity. Second, interviewing our participants about their experiences allows us to understand how the tools were used in accordance with existing routines, in turn contributing to a broader theoretical understanding of collaborative search behaviour 'in the wild'. And, by detailing how our participants used and appropriated particular system features, this research provides implications for the design of future tools to support everyday collaborative search tasks.

6.2 Study Background

6.2.1 Understanding and Supporting Collaborative Search

Within the broader area of *social search*, collaborative web search describes situations "in which participants work together to satisfy an information need" (Morris, 2013, p. 1182). Such activity can occur either synchronously or asynchronously, and participants may be co-located or geographically distributed (Golovchinsky *et al.*, 2008b). Collaborative search is usually scoped to consider explicit, intentional collaboration (cf. Golovchinsky *et al.*, 2011), precluding consideration of filtering or recommendation tools that utilise prior searches from anonymous 'collaborators'.

Several studies have revealed that collaborative search occurs during a variety of professional and personal tasks, including medical research, travel planning, and online shopping (Morris, 2008, 2013). However, searchers report that managing such collaboration can be arduous, requiring workarounds such as link sharing via email (Morris, 2008, 2013) or the use of 'tools-at-hand' (e.g., blogs, text documents) to keep track of search results (Capra *et al.*, 2010). Researchers have suggested that these behaviours can be supported at the user interface, and, to this end, a number of systems have been designed by the research community. Examples include *CoSearch* (Amershi & Morris, 2008) *Results Space* (Capra *et al.*, 2012), *Cerchiamo* (Golovchinsky *et al.*, 2008a), *Querium* (Golovchinsky *et al.*, 2012), *Coagmento* (González-Ibáñez & Shah, 2011), *ViGOR* (Halvey *et al.*, 2010), *SearchTogether* (Morris & Horvitz, 2007), *WeSearch* (Morris *et al.*, 2010), and *CoSense* (Paul & Morris, 2009). While the specific functionality offered by each tool is different, and can vary according to the anticipated scenario of use, the common goal has been to alleviate the need for workarounds by providing browser-based support for collaborative search. Rather than provide an in-depth review of these systems, which is not necessary for the purposes of this thesis, it is instructive to focus on the general aspects of collaboration that researchers have aimed to support with these tools. These are *awareness, division of labour, persistence*, and *sensemaking* (Foley & Smeaton, 2010; Halvey *et al.*, 2010; Morris & Horvitz, 2007; Paul & Morris, 2009). Here we will briefly review the relevance of these concepts for collaborative search and consider how each has been supported in the aforelisted systems.

6.2.1.1 Awareness

Awareness in collaborative search refers to the ability to acquire knowledge about the current and past activities of one's interaction partners (Shah & Marchionini, 2010) which provides a context for one's own activity (Dourish & Bellotti, 1992). Such knowledge promotes coordination and lessens the need for explicit communication about task progress (Gutwin & Greenberg, 2002).[29] Features for supporting awareness during collaborative search include shared query, browsing, and page visitation histories (Capra *et al.*, 2012; González-Ibáñez & Shah, 2011; Morris & Horvitz, 2007); commenting of pages (González-Ibáñez & Shah, 2011; Morris & Horvitz, 2007); and increased salience of particular results based on

[29] An in-depth treatment of awareness in collaborative systems more generally is given in the following chapter.

collaborator ratings (Capra *et al.*, 2012).

6.2.1.2 Division of Labour

Division of Labour in collaborative search refers to the process of distributing a task across members of a group (Foley & Smeaton, 2010). The aim here is to facilitate concurrent work while preventing redundancy and duplication of effort. In collaborative search tools, division of labour has been supported through text chat systems, which allow collaborators to establish division of labour through communication (González-Ibáñez & Shah, 2011); automated splitting of search results (Morris & Horvitz, 2007); allocation of tasks by role (Golovchinsky *et al.*, 2008a); and algorithms for selective filtering (Foley & Smeaton, 2010).

6.2.1.3 Persistence

Persistence, referring to the storage and display of activity from prior search sessions, enables asynchronous collaborative search through information re-finding and resumption of prior search sessions (Capra *et al.*, 2012; Morris & Horvitz, 2007). Persistence has been supported through retention of chat logs, pageview statistics, and session histories (González-Ibáñez & Shah, 2011; Halvey *et al.*, 2010; Morris & Horvitz, 2007); automatic generation of session summaries (Morris & Horvitz, 2007); and relevance rating tools (Capra *et al.*, 2012).

6.2.1.4 Sensemaking

Finally, *sensemaking* support allows collaborative searchers to understand the search process, in terms of *what* has been found, *how* it was found, and where tasks have been handed off between collaborators (Paul & Morris, 2009). Example features include: context awareness through visualisation of search strategies and trajectories (Paul & Morris, 2009); functions to exchange sections of webpages (Morris *et al.*, 2010); and browsable timelines of pages viewed by collaborators (González-Ibáñez & Shah, 2011; Paul & Morris, 2009).

It should be noted that these design aspects are interrelated to an extent. For instance, chat functionality can support both awareness and division of labour by allowing searchers to exchange information and then coordinate their efforts based on a negotiated protocol. The level of provision for each concept may also differ according to anticipated scenarios of use; some systems are directed towards supporting asynchronous search activity (e.g. Capra *et al.*, 2012), whereas others are more generalisable across circumstances (e.g. González-Ibáñez & Shah, 2011).

While studies have suggested that the features described above are beneficial for supporting collaborative search, no work has yet examined the success of any tool when used in natural field settings. Instead, most have been studied using artificial tasks completed under the constraints of short-term laboratory evaluations. This is problematic because real-world search may involve protracted behavioural patterns that are not well supported by current systems. Additionally, a growing body of empirical work suggests that various factors, including awareness (Shah & Marchionini, 2010), communication channel (González-Ibáñez *et al.*, 2013), time (González-Ibáñez *et al.*, 2012), and spatial proximity (Shah & González-Ibáñez, 2010), can impact the collaborative search process. While such studies serve to advance theoretical accounts of collaborative search, they do not offer findings regarding the use of tools outside the laboratory, and offer little in terms exploring how work is organised with respect to division of labour and fairness.

An additional motivation for the present study is that the actual success of collaborative search systems, in terms of mainstream takeup, has been fairly limited. To the best of our knowledge, none of the tools presented within the research literature has achieved widespread adoption. Morris (2013) further notes that the majority of commercial systems intended for collaborative search are either defunct (e.g. *Aardvark, Flock*) or remain in the early stages of development (e.g. *Pinterest, So.cl*). One potential reason for this is that existing collaborative search tools may require too much effort or do not offer meaningful benefit over ad hoc practices (e.g. link sharing via email) used during everyday tasks Morris (2008). However, neither of these possibilities has been investigated. A related issue is that collaborative search tools have only been studied in relatively short-term laboratory evaluations. While such studies are fine for testing specific hypotheses and evaluating initial designs, they do not provide information about how systems are used over the longer term. No work has, as yet, studied the efficacy of collaborative search tools in quotidian settings. It is, therefore, unclear as to how well existing tools fit with everyday collaborative search practices.

Given the considerable research effort invested in supporting collaborative search, we believe that studying how current tools fare in everyday settings would be beneficial for directing future design activities. To this end, we designed a field study of collaborative search tool use, with the aims of gaining a better understanding of collaborative search in the wild and of identifying potential enhancements for future systems. We used two existing systems to study collaborative search behaviour. The fact that we did not develop either system means that we cannot access system logs containing quantitative data about tool use. The study instead focuses on qualitative analysis of interviews conducted after our participants had used an assigned system to complete a real-world collaborative search task.

6.3 Thesis Study 6: A Field Study of Collaborative Search Behaviour and Tool Use

6.3.1 System Choices

We used two systems to study collaborative search behaviour. Our first system was *Coagmento*[30], a freely available tool that incorporates a range of features designed to support collaborative search (e.g. shared bookmarks, chat functionality) (González-Ibáñez & Shah, 2011). This tool can be regarded as a 'general purpose' collaborative search tool in that it is not intended to support any particular task over another.

[30]http://www.coagmento.org

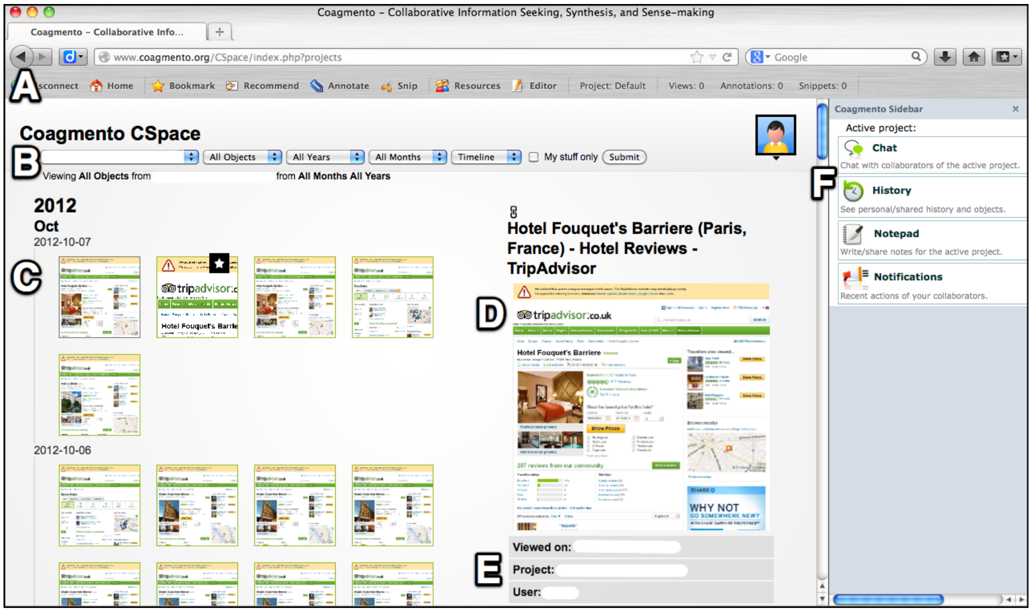

Figure 6.1: The Coagmento system. (A) The Coagmento toolbar. (B) The CSpace with drop down selections for filtering saved results. (C) Shared history displaying thumbnails of recorded pages. (D) A larger preview of a page, accessed by clicking on the relevant thumbnail. (E) Information about a page's time of capture, project, and username of the original viewer. (F) Sidebar providing chat functionality, history, notifications, and a notepad.

Since we did not specify in advance that our participants should engage in any particular type of search, this made Coagmento an appropriate choice for our study.

Our initial intention was to have all of our participants use Coagmento. However, a software update introduced a number of bugs (e.g. SQL database errors, malfunctioning UI elements) halfway through the study. These problems were beyond our control as we are not responsible for the creation of Coagmento. To avoid the problems affecting our remaining participants, we decided to switch to *Diigo*.[31] While this choice was partly pragmatic (very few of the systems presented in the research literature are available for outside use) Diigo's overall functionality is actually very similar to Coagmento, meaning that a change of system did not require a large departure from our established methodology. And, as will be seen in our results, using a second system was beneficial in that differences between the two tools allowed us to obtain some valuable comparative insights.

6.3.1.1 System Functionalities

Both Coagmento (see Figure 6.1) and Diigo (Figure 6.2) are web-based systems. Each is comprised of two parts: a web browser plugin that provides rapid access to features intended for collecting, sharing, and saving information; and an online space to which users can save pages and view the results of prior search sessions. Both tools are also available as mobile applications, but these will not be discussed here as none

[31] http://www.diigo.com

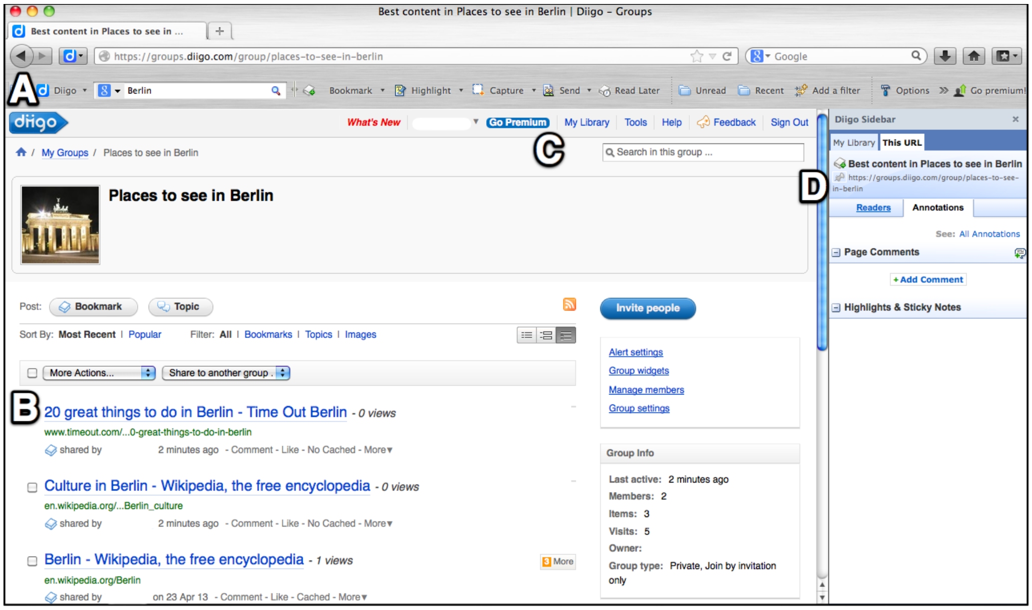

Figure 6.2: The Diigo system. (A) The Diigo toolbar. (B) Shared history with links to recorded pages. (C) Link to the user's personal library ('My Library'). (D) Sidebar providing a list of pages from their personal (not group) library and a list of annotations on the current webpage.

of our participants opted to use them during our study.

Figure 6.1 shows the Coagmento system installed on a user's browser. The browser is open at the online 'CSpace' repository (B) to which pages are saved during search sessions. When the user is logged in to the system, all webpages visited are recorded to this history, appearing as thumbnails in chronological order (C). Each thumbnail can be clicked to reveal a larger preview of the relevant hyperlink (D), alongside the date and time of viewing and the username of the visitor (E). This history is then shared with all collaborators in a dedicated project folder, accessible via the CSpace.

The Coagmento plugin provides two components: a toolbar (Fig. 6.1, A) and a sidebar (F). The former includes buttons that allow users to bookmark whole pages, collect annotations and snippets, and recommend pages to collaborators. Pages captured using these tools are saved to the shared history and appear as an image bearing the relevant icon (the second thumbnail in area C shows the icon for a bookmarked page). Users can then selectively filter their history according to these differing types of content using the 'All Objects' dropdown (B). Of the remaining buttons, 'Resources' opens and closes the Coagmento sidebar, which contains a shared chat, a notepad, a history of recent bookmarks, and a notification submenu. 'Editor' provides access to a shared document for collaborative editing. Finally, the toolbar displays a summary of the current page in terms of views, snippets, and annotations (González-Ibáñez & Shah, 2011).

Diigo (Fig. 6.2) provides a toolbar plugin (A) with functionality similar to Coagmento. Users are able to bookmark pages, leave highlights on relevant sections, or capture (screenshot) particular sections of a page. Each of these can then be saved to the Diigo webspace, either to a private library (accessed via the

'My Library' tab in area C of Fig. 6.2) or, in the case of bookmarks and highlights, to a shared 'group' space listing captured links in chronological order (Fig. 6.2, B). Of the remaining buttons, 'Send' allows users to email a page directly to a collaborator, and 'Read Later' allows for the webpage to be saved to a private library for later reading. The Diigo sidebar (D) provides quick access to the user's private library, a list of annotations for the user's current page, and a view of all prior readers of a particular webpage.

Comparison of Figures 6.1 and 6.2 reveals some differences which are worth considering due to their reference later in the paper. The first concerns the way in which each system saves and presents pages to users. In Coagmento, pages are represented using small thumbnails, each of which must be clicked to reveal further information about the represented page. In Diigo, no thumbnail is visible; instead, the title of the page is shown alongside a hyperlink and the username of the person who made the bookmark. Thus, the provision of initial descriptive information is different in each system—Coagmento relies solely on visual information, whereas Diigo uses text.

A second difference concerns the way in which each system tracks user behaviour. While active, Coagmento records *all* of the pages a user visits while logged in to the system, with individual pages represented by individual screenshots in the CSpace area (Fig. 6.1, C). This means that every page a user visits is captured by the system, regardless of whether or not the content is relevant to the user's primary information need. In contrast, the only pages captured by Diigo are those that the user explicitly tells the system to save, either by bookmark, highlight, or capture (area B of Fig. 6.2 displays some examples). As will be seen in our results, this 'all or nothing' dichotomy leads to a number of concerns related to sensemaking and privacy during the display of shared search histories.

At this point, we should stress that it is not the aim of this chapter to 'evaluate' Coagmento and Diigo in terms of their usability or relative successes and failures. Rather, our use of these systems is guided by the fact that they can be regarded as exemplary tools that might be used to support collaborative search—indeed, Coagmento has been designed in accordance with the research literature on collaborative search (González-Ibáñez & Shah, 2011). We have no investment in either system and no desire to to demonstrate that one is better than the other. Our aim is to use these tools as probes to learn more about real-world collaborative search while identifying broader lessons for future systems.

6.4 Method

6.4.1 Study Design

Beyond our use of two tools, we designed the present study to be as naturalistic as possible. Three concerns here were the selection of search tasks, expected duration of participation, and location of study.

Choice of search task is a difficult yet critical issue when studying information seeking (Kelly, 2009). As we were aiming for high external validity, we elected to use natural tasks (Kelly, 2009) and allowed participants to choose their own tasks. This encourages intrinsic motivation while providing insights into real everyday search behaviour.

The present study was open-ended and no time constraints were imposed on our participants' search process. Not only did this provide ample time for our participants to explore and become familiar with system features, it allowed us to understand tool use during the broader collaborative search process, spanning tentative exploration through to results collection, refinement, and eventual selection.

Lastly, we allowed participants to use the systems according to their own preferences: at home, at work, or even on the move. We felt that learning about where, when, and how collaborative search occurs would help to understand how tools are used in the wild, and might also prove useful in terms of designing future technologies.

6.4.2 Participants

A total of 16 participants (eight pairs) took part in our study. Participants' ages ranged from 19–34 (M = 24.4, SD = 4.7). Pair 3 were both male; pairs 6 and 7 both female; and all other pairs were male-female. Pairs were comprised of friends or romantic partners, i.e. there were no anonymous pairings of unfamiliar participants. Participants were recruited via Facebook and our University noticeboard. Our adverts stated that we were looking for groups of people who would soon be completing a collaborative search task, offering the "opportunity to use a system designed to support collaborative information seeking behaviour". We used purposive sampling, vetting those who responded to our advert to ensure that their information needs were genuine. We did not have cause to turn away any of those who responded to our adverts. We offered each participant £20 as a goodwill gesture for completing the study. All participants were made aware from the outset that their payment was fixed and that it was not related to performance or time spent using their assigned system.

Table 6.1 provides information about the system assigned to each pair alongside task choices. Each pair used only one system and none changed tool during the study. The tasks chosen by our participants fit well with what the literature identifies about collaborative search (Morris, 2008, 2013), and several pairs actually chose to complete more than one task during the study. While the nature of the information required by each of the tasks is different, we consider them to be qualitatively similar in that each is open-ended and allows for considerable latitude in terms of exploratory search behaviour. Furthermore, each involves an evolving information need that calls for the search and comparison of multiple sources with the aim of arriving at an agreed outcome, i.e. a specific location, hotel, or other item of choice. We felt this stylistic similarity would permit generalisation of insights over different tasks.

Table 6.1 also displays statistics regarding distinct search sessions, the number of items captured during each task, and total duration of use. We requested these figures from participants after task completion (this helped prevent advance notice of scrutiny impacting behaviour during searches) and all participants consented. It is worth noting that the total duration of use refers to the elapsed time between participants' first and last search sessions during the entire study. This means that aggregations of completion time for different tasks do not always equate to total duration of participation; some tasks were completed concurrently, whereas others were distinctly separate. Although the lack of precise log data makes these

Pair	System	Chosen task(s)	Distinct search sessions over time, by task	Pages saved, by task	Total time
1	Coagmento	Travel planning	10 sessions over 14 days	27	14 days
2	Coagmento	Travel planning; House hunting	6 sessions over 16 days; 3 sessions over 8 days	10; 8	35 days
3	Coagmento	Concert venues	3 sessions over 12 days	4	12 days
4	Coagmento	Travel planning	3 sessions over 7 days	9	7 days
5	Diigo	Houses; Shopping	3 sessions over 5 days; 8 sessions over 3 days	3; 10	10 days
6	Diigo	Houses; Travel planning; Shopping	15 sessions over 14 days; 4 over 7; 3 over 3	18; 8; 14	21 days
7	Diigo	Houses; Travel planning	3 sessions over 7 days; 2 sessions over 8 days	22; 8	21 days
8	Diigo	Travel planning; Shopping	8 sessions over 6 days; 5 sessions over 8 days	10; 8	14 days

Table 6.1: Participants' assigned system, task choices, distinct search sessions, number of items saved, and total duration of use. The 'Pages saved, by task' column refers to the total number of items captured using bookmarks, snippets, or annotations. Note that some tasks were completed concurrently and others were distinctly separate, meaning that aggregations of task completion do not equate to total duration of participation.

somewhat coarse indicators of usage, it is clear that all participants used their assigned system for at least one week and engaged in multiple information seeking episodes during that time. This in turn increases our confidence in the meaningfulness of participants' experiences with their assigned system.

6.4.3 Materials & Procedure

Each pair of participants was provided with all necessary software alongside installation instructions and a briefing script introducing the study, tailored according to the system assigned to each pair. The script clarified that participants should use the software we had provided each time they were searching for information towards their chosen task. The script also stated that there were no expectations about the way in which the tool should be used. Rather, participants were encouraged to use their tool in whatever way they deemed appropriate. We also provided a detailed instruction document explaining the functionality of the relevant system. The document described each system feature in full by providing a screenshot alongside explanatory text. All features were given equal descriptive treatment (roughly one page per item) so as not to suggest the importance of any feature over another. We were also careful to ensure that we only described the workings of each feature, rather than how it should be used to support search behaviour.

After installing the relevant system, participants created their own user accounts and were allowed to search at their leisure. As mentioned above, no constraints were placed on participants in terms of process. Participants were free to decide when to engage in search sessions, how long to spend on each session, and when to terminate their information seeking activities. Participants were asked to email the author once they felt they had satisfied their information need and no longer had use for the system. A semi-structured interview was then performed with each participant. Thirteen of these interviews were conducted face-to-face in a quiet office; two were over the telephone; and one was via Skype. Interviews were one-to-one between participant and the author; searchers did not, therefore, discuss their information seeking behaviours while their partners were present. This allowed us to cross-check statements and ensure consistency of behavioural accounts. During interviews, participants were invited to access their assigned system if they could not remember exact details or wanted to elaborate on particular functionalities.

All interviews lasted less than one hour. To avoid response bias and to dissuade participants from trying to please us with their answers (Dell *et al.*, 2012), we began by reminding participants that we did not design the system and there were no right or wrong answers insofar as this study was concerned. We used a basic framework of 32 questions, as listed in Appendix G of this thesis, which were directed by the concerns of our study and questions used in prior work on collaborative search (Capra *et al.*, 2010). These questions allowed us to explore participants' search process; interactions with their tool and collaborator; and how search products were used in accordance with their chosen tasks. As our protocol was semi-structured, we were able to probe issues as and when they arose, and we invited participants to elaborate on their remarks, allowing novel topics to emerge.

6.4.4 Analysis

All interviews were conducted, recorded, and transcribed by the author of this thesis. This procedure was adopted in favour of automated or paid transcription services because transcription is integral to the process of analysis; by coding the interview transcripts himself, the author was familiarised with the data at an early stage (Riessman, 1993) and was able to interpret meaning conveyed by participants' intonation and delivery (Braun & Clarke, 2006). This fed through into the transcription of interviews, which were transcribed so as to give a verbatim account of participants' statements for later interpretation (in line with recommendations in relevant literature, e.g. Lapadat & Lindsay, 1999; Braun & Clarke, 2006). This resulted in over 200 pages of transcript.

We used open, axial, and selective coding (Strauss & Corbin, 1998) to identify initial codes, structure codes into themes, and then group themes with reference to participants' search process and behaviour. The coding process was iterative, with transcripts subjected to three readings. Initial codes were developed and refined during each reading, with internal consistency achieved by scrutinising the data for counter-examples. We should acknowledge that some codes were influenced by our interview questions. For example, by asking about division of labour, it was inevitable that a related code would arise during analysis. However, many new codes also emerged, indicating to us that the results had novelty beyond current knowledge.

6.5 Results

To bring clarity to our results while making sense of our participants' experiences, we cluster our themes using four categories: *search process & management*, *appropriations & afforded behaviours*, *fairness preferences*, and *interface design issues*. The first encapsulates general information about our participants' collaborative search process; such insights are detailed independently of specifics concerning how system features were used. The second pertains to the emergent forms of interaction centred around the use and appropriation of particular system features in support of collaborative search. These categories address our aim of exploring real-world collaborative search. The third considers the extent to which participants were concerned about fairness during task assignment and execution. The final category allows us to drill down into specific issues linked to interface design, many of which are best understood through consideration of participants' search behaviours.

As our data are primarily qualitative, we present themes alongside direct quotations from our participants. To identify speakers, we use the form [P*x,y*] where *x* indicates the ID number of the pair, as listed in Table 6.1, and *y* refers to the first or second member of that pair.

It is worth noting that participants were generally positive about the design concepts behind the two tools, and all participants believed that their assigned system had merit over ad hoc solutions. Regarding use of system features, all participants stated that they used their system's toolbar, and that they had accessed the shared online space to view search results. In Coagmento, searchers reported using bookmarks, annotations, and snippets. All participants stated that they did not use recommendations, the shared editor, or any of the resources in the sidebar. In Diigo, participants reported using highlights, bookmarks, and screenshots. All participants stated that they did not use Send, Read Later, or any of the functionality within the sidebar. While the absence of log data means that we cannot definitively state that particular features were or were not used during our study, participants' statements are revealing about preferences for some features over others. Before providing more specific details about how features were used to support collaboration, the following subsection paints a general picture of participants' search process and management strategies.

6.5.1 Search Process & Management

6.5.1.1 Circumstances of Search

Table 6.2 outlines the search scenarios reported by each pair of participants. Although we cannot specify exactly how many times each type of search occurred, we did ask participants which of their identified scenarios occurred most often. Responses indicated that distributed, asynchronous search was the most frequent scenario for pairs 1, 2, 3, 5, and 6. Pairs 4, 7, and 8 identified co-located, synchronous search as their most common scenario. While one might expect that such choices would be determined by living arrangements, this was not always the case. Pair 2, for example, were a cohabiting couple who reported conducting all of their searches asynchronously.

Perhaps a more valuable aspect of these results is the fact that tool use was not isolated to a single setting. Rather, all pairs reported engaging in at least two different scenarios of search, and participants described how search activities shifted according to everyday life. Pair 1, for example, reported working only asynchronously when apart, but engaged in both synchronous and asynchronous search sessions when co-located at the weekends. Our participants did not report any immediate difficulties in managing these transitions; this may be due to the fact that, as identified earlier, both tools used in our study are general purpose systems intended to support search across a range of scenarios.

	Co-located		Distributed	
Pair ID	Synchronous	Asynchronous	Synchronous	Asynchronous
1	✓	✓		✓
2		✓		✓
3		✓		✓
4	✓			✓
5			✓	✓
6	✓			✓
7	✓			✓
8	✓	✓		✓

Table 6.2: Participants' reported circumstances of search activity. A ✓in a cell indicates that participants engaged in the relevant type of search.

6.5.1.2 Task Management Strategies

Four pairs reported using division of labour as a way of organising their task. Pairs 1 and 2 divided their travel planning by determining a set of locations and splitting them to create targeted individual searches. Pair 3 established a very distinct role assignment, where different aspects of the search process were divided using semi-formal roles: one searcher was responsible for finding a venue for their band's concert, while another sought places to promote the event. Finally, pair 7 reported dividing the web space during house hunting, with one searcher checking private rental sites and another scouring university-hosted pages. These strategies reflect variations of the *divide-and-conquer* approach described by Morris (2008) and align well with the strategies identified in Chapter 5.

The remaining four pairs did not plan a division of labour. These pairs described a 'dive in and do it' approach, akin to the *brute force* management strategy identified by Morris (2008). However, participants did not report any negative experiences related to organising the task in this way.

6.5.1.3 Search Process

While we were not able to log behavioural data about search activity, participants did provide verbal accounts of their search process, and these are instructive in terms of understanding what searchers were actually doing during our study. All participants reported following a broadly similar process, beginning with the use of familiar search engines to learn about the information space at a high level. Participants then reported narrowing their search focus to be more specific:

> *"We started by simply Googling what we're looking for, and when you get leads you search for those specific things, like that specific realtor or a specific area."* [P6, 2]

Participants reported that this process was directed by their existing knowledge, and by external factors such as an available shopping budget, dates of travel, or availability of particular venues in the local area:

> *"I already knew which websites to target, I started off with a few commercial websites and others, looking at StudentPad and the [University] noticeboard."* [P6, 1]

After identifying relevant websites, behaviour shifted to the selective capture and bookmarking of results. For all our participants, search and bookmarking were fragmented over multiple sessions. Searchers stated that sessions were interspersed with communication, both during and between searches. As well as ordinary conversation, email, and message exchanges, participants reported using system features like comments and annotations to further their task-related discussions. Participants reported that their overall tasks would end with discussion of the results to reach consensus over an outcome, i.e. the 'chosen products' from their search. In broad outline, this process corresponds well with existing theoretical accounts of web search, where exploratory information seeking has been characterised as an extended, fragmented process, involving multiple search sessions, queries, and information sources (Golovchinsky *et al*., 2012).

We now delve into the way in which system features were adopted in support of this general pattern of collaborative search behaviour. Our later Discussion section then considers the broader implications of the results, in terms of lessons for collaborative search systems.

6.5.2 Appropriations & Afforded Behaviours

6.5.2.1 Page Capture for Re-Access and Suggesting Relevance

During their information seeking process, all eight pairs made use of system features intended for information capture: bookmarks, snippets, and annotations in Coagmento; and bookmarks and highlights in Diigo. At the most basic level, these features were used to overcome the ad hoc practice of sharing hyperlinks via email. Participants valued the ability to save pages to a joint repository and, correspondingly, to see pages saved by their collaborator:

> *"I thought it was really helpful to know exactly what he had looked at and what he wanted to go see. Instead of him having to send me a variety of different emails and links, it's all been saved visually right there for you."* [P4, 1]

In providing different features for capturing information, the designers of both systems have anticipated that searchers may wish to retain different types of content. For instance, bookmarks are intended as a way of saving whole pages, whereas annotations and snippets allow capture of particular subsections or page elements. Our participants, however, tended to adopt just one of these features to achieve the same outcome: that of capturing pages to 'pull out' particular results from the web space. Pair, 1, for example, used Coagmento's 'Snippet' feature to achieve this, whereas pairs 2, 3 and 4 used 'Bookmark' for the same purpose. Furthermore, responses indicated that capture behaviour was more nuanced than simply 'saving pages for later'. First, capture was intended as a way of favouring particular results for more rapid re-access:

> *"It was particularly useful because I could go back quite easily and know what I'd already put up there, so I wasn't reposting it or covering the same ground."* [P5, 2]

Second, capture was sometimes used as way of increasing the salience of particular results, with the aim of suggesting potential relevance to a partner. The intent was not necessarily to return to a particular page but was instead an attempt at bringing something to the partner's attention:

> *"Looking through Expedia, when I found some good hotels I would snippet them and then I would say to him, oh I've found these ones that look good, and I've snippeted them."* [P1,1]

6.5.2.2 Forming and Discussing Shortlists

We identified that page capture was actually related to a more general strategy—that of *shortlisting*, where features like bookmarks and snippets were used to form lists of potential choices and, eventually, reach an agreed outcome. All participants engaged in this behaviour:

> *"We would aim towards a list of five nice hotels"* [P2, 2]

> *"In the final decision, we had a top five possibles to choose. Which is where the bookmarking came in."* [P4,2]

Participants testified that shortlisting was an existing behaviour, but espoused the benefits of the tools in terms of bringing structure to this process:

> *"Normally we would put links on Facebook, drop all the links into a conversation. But that is chaotic. So it was nice to have them all there without going through our conversation trying to find the link we're talking about."* [P7,1]

Participants reported engaging in conversation about their shortlists, either verbally or, in the case of Diigo, by annotating shortlisted links using comments. This occurred as a back and forth process of contributing and evaluating suggestions to the list. Participants stated that using comments on Diigo was useful in supporting this behaviour, indicating the value of features that allow discussion of shortlisted items:

> *"We would keep commenting on each other's things, so one link had three comments, I would comment and she would comment, then I would comment back. And they weren't all mixed up."* [P6,1]

While the actual content of the lists was qualitatively different between tasks (i.e. some pairs were choosing houses, others shopping items) the overall goal was the same: form a list of candidates, discuss and refine the list, and then settle on a 'good enough' outcome in line with the goal of the information seeking process. However, this process was not always linear or in sequence. Some participants reported that they would be more careful about relevance by evaluating sources on-the-go, keeping only 'definite' possibilities in the shortlist. Others would take a more carefree approach, forming a longer list of potentially relevant sources and then narrowing them down at a later stage. Participants reported that shortlists were eventually used and reviewed to reach final consensus:

> *"A lot of it was down to seeing what the other person thought, and then between the two of us, we decide, oh let's go for that one. If there's no availability there, you go to the next choice on your list. So you almost rank them between the two of you, for personal preference."* [P1,2]

6.5.2.3 Sharing versus Saving: Sensitivity to Time Constraints

In forming shortlists, participants reported that if their task was time pressured and needed to be completed in a short time frame, they would change their behaviour to ensure that their partner checked recorded links more quickly. In other words, rather than wait for their partner to check the system at some unspecified time, participants would use external communication methods to alert their partner about the information:

> *"We want a house quickly because it's a little late, so we had to speed up the process. I had to call her up and tell her ok, please go and check this house, I've posted it on the group. It's possible she would go and check, but we were in a hurry, so I have to convey the message to her ASAP."* [P6, 1]

Participants also explained that information was itself often time-critical, in that pages could be subject to expiry. For instance, holiday deals and special offers can be available for limited periods of time, and houses can disappear quickly from the rental market. The knock-on impact for collaboration was one of requiring fast action:

> *"If it comes up with a great hotel price you click on more details and it would say, there's only two more rooms with this price left... if there's only two rooms left, we need to be quick otherwise we might not be able to get this room at this price."* [P1, 2]

However, participants spoke of using out-of-band channels to notify their partner more quickly in this instance:

> *"Because it gets updated quite often, London seems like quite a quick turnaround on properties. So we chatted over Facebook a bit and then I could send links to websites over Facebook and say does this look ok?"* [P2, 1]

This behaviour is interesting given that both systems have features allowing webpages to be sent directly to the email inbox of a collaborator, both of which might be used in a time-pressured search scenario. Yet none of our participants used these features; instead, ad hoc methods of communication were preferred.

6.5.3 Fairness Preferences

In addition to studying participants' search process and tool use, we asked questions about fairness and whether or not searchers were concerned about equity in the work process. The first thing to note here is that a majority of pairs (six out of eight) believed that their work had been completed roughly equally. While it is important to remember that these results are based on retrospective self reports, most were satisfied about this way of completing work:

> *"I think we both did the same, which was nice. I was satisfied with that."* [P6, 1]

However, when probed further, the majority of participants were actually not all that concerned about fairness during their completion of the collaborative search task. This may initially seem contrary to our contentions regarding the importance of fairness in collaborative work, yet close reading of the responses reveals new and nuanced perceptions that have implications for our understanding of fairness in a collaborative setting. First, participants were working with friends and family and so were happy to let things slide that might not be forgiven in professional settings. Such responses suggest that the context of work influences the salience of fairness norms:

> *"We're friends, so one it doesn't count... it doesn't matter if I put in two and a half hours and five links and she puts in an hour and a half and fifteen links, as long as we find it."* [P6, 2]

> *"Fairness wouldn't be important to me, no. If it had been someone I didn't know, something like you've actually got work to do, if it had been uneven that would have been annoying."* [P5, 1]

Second, the nature of the work itself seems important as many participants did not see the task as 'work' *per se*. This was true of the task itself and the consequences that would stem from failing to contribute:

> [Fairness didn't matter] *"because it wasn't like 'work', it was fine... But if you're both meant to be contributing equal parts to a job, and somebody else is not pulling their weight, it's annoying... but here, there's no real implication of someone not doing it. There's no kind of repercussions or anything."* [P8, 1]

Two individuals, one from pair 3 and one from pair 4, believed that their work was not completed fairly. However, contrary to our earlier study of student workgroups, respondents were not all that concerned about the apparent lack of fairness. These participants felt it was more important to ensure task completion over equity:

> *"I think that getting the task done was more important."* [P4, 1]

> *"We needed to get it done, so just having her quick opinions was worth the workload being uneven."* [P5, 1]

In some cases it seemed like ensuring that both had opportunities to participate, regardless of how much work was done, was more important than ensuring equality. This was related to a need to ensure that both members were involved simply so that they each had the opportunity to give input on the products of the collaboration:

> *"To a certain extant, I wanted to make sure that it was being equally done, and that it wasn't just me doing it while someone else played Sim City, because when we're travelling further down the line all of a sudden someone's like, why are we staying at this hostel?"* [P4, 1]

However, a key issue arose in that, while most participants were not that concerned about fairness, some said that it was actually quite hard to make judgements about what their partners had been doing. This issue ties in with some of our earlier contentions regarding the difficulty of judging contributions (and hence, equity) in computer-mediated settings:

> *"It was hard to tell what he had been doing and how much he'd done without seeing his screen."* [P4, 1]

The fact that searchers found it hard to gauge contributions during use of these two tools is somewhat surprising given that both systems have a shared search history. Each of these could presumably be used to make judgements about fairness—one could simply visit the history and compare the number of sources accessed to work out whether or not contributions were even. However, as will be seen in the next section, participants encountered various issues with this history, many of which caused them to forego using it altogether. Thus, in truth, it is likely that participants were left unable to make considered judgements about fairness—this is a concern that has been raised in this thesis and elsewhere in the literature (e.g. Galegher & Kraut, 1990). For now we shall defer this matter to our discussion section. The following subsection delves into interface issues encountered by our participants.

6.5.4 Interface Design Issues

In addition to the behavioural patterns identified above, participants described a number of experiences and issues related to interface design and collaborative search.

6.5.4.1 Shared Search Histories: Information Overload

Participants described how, during search sessions, they would visit the online repository of their assigned system to review their own results and explore what had been found by their partner. Recall that our first system, Coagmento, provides a complete history of search behaviour by capturing all pages visited by the user during information seeking episodes. Previous work suggests that such 'search trajectories' can be beneficial for sensemaking and facilitating awareness (Paul & Morris, 2009, 2011). However, seven out of eight Coagmento users actually reported that they found the history overwhelming:

> *"I'd open up the hostel search and there'd be a list of a hundred hostels... and when I'd try and go back to look at which ones were the cheapest, I'd have to look through all of them again to try and find the one in the history. I ended up just searching it again because it was too hard to find."* [P4, 2]

This problem was partly related to the collection of landing pages, portals, and otherwise irrelevant material that held little utility for sensemaking. Participants suggested that they wanted to get rid of such pages and see only the most relevant information without having to trawl through the entire history:

> *"I never looked at the history because the snippets were the core information. I just want something that jumps out and says... this is the information that she has found."* [P1, 2]

Participants suggested that it would be helpful if Coagmento's history was searchable, or that, rather than allowing the system to indiscriminately record all pages, the process of information capture could be more selective:

> *"It's almost like when you're searching you want to be able to dictate what's saved. So even if it's just a little 'plus' button, where maybe you don't want your Google search to come up, but, whatever site, even if you don't think it has anything to do with it, but you might want to look at it again, you can just hit a button and it's going to come up there."* [P4, 2]

Some users did mention benefits of the timeline view. One member of pair 1 reported retracing their partner's search results to see what had been done earlier. Members of pair 2 reported using the timeline to direct their own work based on the results of their partner, while one member of pair 4 thought the ability to rediscover past results could be useful.

Responses from Diigo users provide an interesting contrast on this issue. Such participants did not express any dissatisfaction about not being able to view a complete search history, and none expressed any desire to have more information about their partner's search process. Some were actually skeptical about the need to understand their partner's process:

> *"As long as you both know what you're looking for, it doesn't matter how you go about finding it. I didn't feel like I needed to know what he typed into Google."* [P8, 1]

6.5.4.2 Shared Search Histories: Privacy Concerns

The fact that Coagmento captures images of all pages visited made three users hesitant about the potential consequences of having personal information recorded:

> *"While I'm searching for hostels I'm using Facebook and Gmail at the same time... I think if you're working in a group that history could get a little weird, because you're going back to your bank account to figure out how much you still have left."* [P4, 2]

This highlights how, for our participants, real-world collaborative search was not an isolated endeavour; instead, it occurred as part of a broader planning activity that required access to personal and private information. Clearly a screenshot image of such content would risk a privacy violation. Participants were not specific about whether any such violations occurred during the study, but the mere threat of information leakage was enough to affect behaviour. In line with earlier statements concerning information overload, participants suggested the solution of selective tracking:

> *"I was always making sure that I wasn't logging in to other things while logged in to that. Maybe if it had a way of knowing it was on a page for logging in to your emails... it would know not to record those kinds of pages."* [P3, 2]

In contrast, privacy did not arise as an issue during discussion with Diigo users, likely due to the fact that no pages are captured unless specifically requested by the user.

6.5.4.3 Shared Search Histories: Sensemaking and Rationale

Several issues arose related to sensemaking of the shared history, in terms of understanding *what* and *why* information had been found. Regarding the former, responses suggest that searchers need to be able to appraise page representations both quickly and easily, and do not want to spend a long time making sense of shared results. This was exemplified by participants' experiences with Coagmento, where each item in the shared history appears as a small thumbnail showing an image of the webpage (see Figure 6.1). However, rather than simple thumbnail images, participants wanted an up-front summary of contextual information about each page: a descriptive headline, a preview of its contents, and the time it was saved. We note that this contextual information is available within Coagmento but requires the user to click on each thumbnail to view it. Our responses imply that such information should be presented up-front, with low interaction costs, to facilitate rapid assessment of relevance and sensemaking of pages:

> *I would want to see the header you'd get if you bookmarked it in a browser. So, under 'hotel search', then the name of the hotel, that sort of thing. The thing is that it's either two or three links you have to click through before you even get to the page itself...And so I would've used that more if I'd had the ability to go, I know what that is... that's what I want... without having to click, click, click."* [P2, 2]

Although Diigo records pages using a title and contextual information, sites are not listed with thumbnail images (see Figure 6.2). Somewhat ironically, all Diigo users indicated a desire for thumbnail previews of each page, as with the style used by Coagmento. This suggests providing both types of information could be beneficial for sensemaking:

> *"You know like on Facebook when you put a link they give you a small image and a description of what's in the website... it would be nice to have that feature because it triggers your mind about the thing you saw... a picture of the house could appear there automatically."* [P7, 1]

A second issue related to sensemaking was that searchers wanted to understand *why* particular pages had been visited by their partners. This was true of both the search history and pages shortlisted using bookmarks:

> *Without the other person telling you what they'd gleaned from each of the links, it was difficult. You need some explanation, he could have looked at all this stuff and thought, this is a load of rubbish. Then what's the point of you looking at it?* [P1, 1]

Participants wanted to annotate and append specific pages to provide rationale about why results had been selected, and to draw attention to specific aspects of webpages. Both practices are in line with the act of suggesting relevance:

> *"When I would bookmark a whole page, it wasn't necessarily the whole page that I wanted her to look at... I wanted her to pick out the Regent's Canal, but if I'd bookmarked the page, she wouldn't necessarily know that."* [P4, 2]

Users of Diigo reported using a short description at the point of page capture to inform their partners about why particular links were relevant. This information would then appear alongside each result in the shared space. Searchers also described appending links with information gleaned from the page so that their partner did not necessarily have to re-run the search. However, users also mentioned that they wanted more ways of tagging or marking captured results. One goal was to indicate that certain pages were temporarily irrelevant. Users did not want to delete such pages in case they became useful later on:

> *"I added 'TAKEN' to the house because I didn't want to delete it, it's better to have it there for future reference because you never know, so I just put that tag."* [P7, 2]

6.5.4.4 Segregation and Manipulation of Shared Information

Our chosen systems allowed shared results to be viewed in two forms: either as a list of one's own results, or as a combined list of with those of one's partner. However, four participants desired to see *only the results of their partner*. This action was not possible in either system, but this type of separation would clearly be advantageous when attempting to make sense of a partner's work:

> *"It came up with all of your stuff to start with, what you'd done when you go on it... which I really didn't care about as much, because you know what you've done. That's not the immediate reason you go on there, you go on to see what the other person has done."* [P1, 1]

Participants also wanted the ability to distinguish different aspects of their work from others. For example, pages related to flights, hotels, and sightseeing were treated equally by the systems, i.e. as superficially equivalent 'pages'. But to searchers, these pages were related to different subtasks. Users wanted more ways to differentiate and classify these results by creating subcategories and folders that would allow results to be directed towards particular sections. Only one of our tools, Diigo, provided support for this. Participants were able to create specific groups to manage the overlap in concurrent completion of different search tasks:

> *"We made different groups, one was for house hunting, one was for holiday hunting, and the third one was on these dresses that we were shopping for online."* [P5, 1]

Additionally, users reported that they wanted the ability to manipulate the results in greater depth; for example, by moving sources from project to project in accordance with changes in information need and the overall progress of task completion. Users of both systems also wanted to rank and reorder results in their shortlists, so as to allow for visual comparisons. This related to the practice of shortlisting and reaching consensus:

> *"if I have a preferred order for all these houses, like if I want to call this one first and that one second, I would like to have the possibility of rearranging them."* [P7, 2]

6.5.4.5 Awareness and Notifications

In discussing their partner's activities, participants' responses imply that notifications of recent work (e.g. searches, bookmarks) would help to maintain a general level of awareness:

> *I think it would be nice to get actual notifications of things. If she'd bookmarked a page, I could get a notification of a bookmark, or a tag, something like that.* [P4, 2]

Interestingly, both systems do provide notifications but in different ways. Coagmento's notifications appear in the sidebar but, as all of our participants stated they did not use this feature, likely went unnoticed. By contrast, Diigo provides daily email notifications alerting collaborators to the presence of new links in shared groups. Although one participant found the emails useful in terms of being able to check recent links with her mobile device, the remaining participants said they did not actually read the emails; instead, they were mainly used to gain awareness of the mere fact that *something* had been done. Again, this speaks to a desire for some general level of awareness about the fact that contributions are being made. When checking their assigned system, participants wanted to be able to find these contributions quickly and easily:

> *"It could highlight that this is the new comment, or this is the new thing, instead of me having to search constantly."* [P6, 2]

Two users stated that it would be helpful to have awareness of when their saved results had been viewed by their partner, i.e. a confirmation that results had actually been seen. Responses also indicated a desire for such notifications to be more immediate and situated in the web browser, rather than the system itself. Several drew on their experiences with Facebook to suggest how these notifications might be implemented:

> *"It could work like Facebook, where you get the notification, you click on it and it takes you to the page."* [P4, 2]

6.5.4.6 Effort Requirements

Finally, throughout our dataset, there was a general undercurrent of wanting a minimal threshold for effort. The perception was that, when search was the primary task, any additional effort above and beyond an ad hoc solution was undesirable:

> *"All you wanted was something easy, you didn't want something that was going to add a load of time, because the search was quite time intensive. You didn't wanna feel like you were going to have to do loads of extra work on it."* [P1, 1]

Users of Coagmento made more specific remarks about effort, in that that they would prefer aspects of the tool to have lower interaction costs. In addition to the earlier statements concerning the number of operations required to access contextual information in the search history, participants remarked upon the excessive number of stages involved in other tasks. For example, when the system asks users to rate and leave comments about bookmarks at the point of page capture:

> *"I guess the reason the bookmark thing asks for a one to five rating is so it could do some filtering. But at the time it's too much effort. I don't want to do that, I want to automatically know what's important."*
> [P2, 1]

6.6 Discussion

The present study sought to investigate collaborative search behaviour in everyday settings. At a broad level, the fact that searchers found merit in their assigned system suggests that the design concepts embodied by each are on the right track—it was not the case that participants considered the tools unusable or inappropriate, and many stated informally that they planned to use the tools again outside of our study. Some of the primary benefits noted by our participants included the ability to save links to a shared space, thereby negating the need for ad hoc workarounds, and the fact that the tools brought structure and persistence to the otherwise ephemeral process of collaboration over time.

6.6.1 Implications for Collaborative Search Tools

General aspects of participants' search behaviour are revealing about how the tools were used in everyday circumstances. Beginning with the search process, collaboration among our participants was not isolated to a single configuration of time and space. Instead, all participants reported engaging in multiple scenarios. Although no work has argued otherwise, these findings suggest that tools should aim to support search across a range of settings rather than a single time/space configuration. The systems we used were not restrictive in this regard, but some other tools (e.g. Capra *et al*., 2012; Halvey *et al*., 2010) have been designed with specific circumstances in mind.

One of our most consistent behavioural findings was the generalised use of page capture features (snippets, bookmarks, and annotations). While this stresses the importance of allowing searchers to save pages to a shared space, our results suggest that capturing pages is actually more nuanced than simply 'saving pages for later'. Rather, searchers used these features to save pages for re-access; to share results with their partner with the aim of suggesting relevance; and to form *shortlists*, i.e. subsets of results that are collected as potential candidate outcomes for the group's collaboration. However, we saw that pairs tended to settle on, and persist with, a single means of page capture to achieve all three (e.g. Pair 1 used only snippets for shortlisting and suggesting relevance). Given that these are essentially distinct behaviours, which might appeal to different users in the light of different nuanced needs, we suggest that each is potentially worthy of independent support. Future tools could provide separate methods for *page saving*, *page sharing*, and *suggesting relevance*. Regarding the latter, some existing tools allow collaborators to suggest relevance by 'liking' pages (Golovchinsky *et al*., 2012) or through up- and down-voting of results (Capra *et al*., 2012), but future systems could do more to separate shortlisted pages from those that are awaiting judgements of relevance. This would avoid conflation of the two, as occurred for some of the participants in our study.

In forming shortlists, participants wanted to annotate saved results to share rationale about *why* pages had been selected; to make pertinent information salient (e.g. the rent of a saved property or the price of a hotel); and to engage in task-related discussion about the quality or relevance of results. Several systems (Morris & Horvitz, 2007; Paul & Morris, 2009) allow commenting of saved items in the manner offered by Diigo; our study reiterates the value of allowing annotation of links to facilitate sensemaking and sharing of knowledge in the context of everyday tasks.

However, we found that searchers wanted to rearrange listed items to express an order of preference, or 'check off' items by deleting some while retaining others. This process of bringing structure to data is part of sensemaking (Tao & Tombros, 2013) but our chosen systems did not allow users to repurpose results in any meaningful fashion at the end of their tasks. The implication here is that, as well as supporting aspects of the search process, future tools might also do more to support the manipulation and reuse of search *products*. Since searchers want more ways of interacting with their results, different workspaces could be provided to allow segregation of items; searchers could then be allowed to move items between these spaces in accordance with relevance judgements and task progress. Such an approach can be seen in the ViGOR system (Halvey *et al.*, 2010), which includes a workspace that allows searchers to drag and drop results into groups that can then be reorganised and restructured at will.

Some broader issues related to sensemaking and awareness concerned searchers' difficulty in understanding *what* had been found during their partner's searches. This is a foremost challenge during collaborative search (Paul & Morris, 2009) and indeed this was true for our participants. An initial issue concerned identifying the location of recent work when checking their assigned system—searchers often found it hard to identify recent collaborative contributions (e.g. new annotations) in the presence of a large search history. Future tools can benefit by ensuring that notifications are available and that they lead straight to recent contributions. A second lesson is that sufficient information must be given to allow searchers to appraise shared representations with relative ease. Prior laboratory studies suggest that rapid sensemaking can be promoted by displaying contextual information about search results (Paul & Morris, 2009, 2011). Our participants' statements align with this, and indicate that representations should combine content previews (e.g. a visual thumbnail) with contextual information (e.g. a page description) that allow comprehension while lessening the need for collaborators to click through and revisit every individual result.

An important design challenge raised by the present study concerns the vexed question of history logging and the extent to which page capture should be deliberate or automatic. Earlier work suggests that the approach of capturing all pages, which provides a persistent account of a collaborator's 'search trajectory', is beneficial for sensemaking (Paul & Morris, 2009, 2011). However, we found that, with this approach, the quantity of information captured was regarded as overwhelming. Rather than wade through a large history, our searchers simply wanted to identify only the most relevant information. This finding is similar to that obtained in a laboratory evaluation of CoSense (Paul & Morris, 2011), where searchers reported feeling overwhelmed by search histories and struggled to identify 'good' from 'bad' information. In suggesting solutions to this problem, Paul & Morris (2011) recommended that designers allow filtering of search histories by content. This solution would not, however, necessarily overcome the problem of information overload as pages accumulate over time—even filtered lists could become difficult to comprehend. One interesting aspect of our results is that none of our Diigo users felt the need for a complete search history. This tentatively suggests that a complete history may not always be necessary for quotidian tasks, especially if collaborators communicate about their progress as part of their daily routines. (As in our study.) That Diigo users did not complain about the effort of saving pages also hints that page capture could be left entirely under the control of the user. This would, in turn, prevent the collection of interstitial pages that appear to offer little utility for sensemaking.

This work has also identified issues that arise only in real-world search. One example pertained to time-criticality, both of search processes and information itself, with some items available only for short windows. We are not aware of any prior work that has raised time constraint as an issue for collaborative search—perhaps future systems could elaborate on this issue and provide more pointed ways of delivering results to one's partner (e.g. through mobile devices). A second issue was the subject of privacy, which highlights the need to ensure that tools are capable of interleaving with the broader tasks in which collaborative searches are embedded. Searchers in our study were nervous about the potential capture of personal or private material. While an obvious answer to this issue would be to allow deletion of pages from the system's history, a more elegant solution would be to record only pages relevant to the user's information need. This could be achieved by allowing the user to enable or suppress tracking in specific browser tabs—this would compartmentalise search to a specific location, leaving other areas safe for private multitasking.

As many of these implications concern specific issues, it is worth considering how the present study can direct future collaborative search tools more generally. First, our results lend support to Morris' recent assertion that collaborative search solutions must be low-effort and "sufficiently lightweight compared with status quo ad hoc solutions" (Morris, 2013, p. 1190). Our participants expressed dissatisfaction if interactional demands imposed by their tool exceeded those of their previous solution (e.g. email, Facebook). Since the implication here is that any tool with unnecessary effort requirements stands to fail, features in collaborative search systems should be benchmarked against the equivalent ad hoc solution. Additionally, the tools used in our study included many features which, according to our participants, were not necessary for their task. This suggests that future solutions could be scaled back in favour of lightweight support for core collaborative search behaviours. An example system might support the rapid sharing of pages between two linked browsers, with simple awareness mechanisms, like those desired by our participants, that notify collaborators of recent activity after their web browser is opened.

Alternatively, more specialised systems could be developed to provide targeted support for specific tasks. Some of our participants mentioned informally that they would value support beyond that for search. In travel planning, for example, searchers could be able to apply their shortlists to a map, helping them with their journey or sightseeing plans. The wider implication here is that collaborative search tools could be embedded in larger applications that support a broader range of high-level planning tasks.

6.6.2 Division of Labour and Fairness

Regarding division of labour, our first result was that the types of coordination strategies used by participants in our earlier laboratory studies were replicated in the field. One pair divided the work by partitioning the Web space. Two others assigned labour by splitting travel searches according to potential destinations, very much akin to a division by semantic space. Additionally, one pair used a role assignment. This latter strategy extends our earlier work and emphasises the utility of distinct roles for division of labour (e.g. Golovchinsky *et al.*, 2008a). Other pairs did not adopt formal divisions of labour and instead used a brute-force strategy (Morris, 2008). Since pairs did not report any difficulties with managing their task, we can only assume that these approaches were sufficient for managing work—perhaps searchers would have encountered problems if aspects of their tasks were more tightly coupled. Moreover, it appears that work was completed asynchronously in a majority of cases, meaning that the risk of redundancy was easier to overcome. Participants reported discussing their findings daily, and it is not difficult to imagine that these discussions would direct search in directions that differ from those already explored by the partner.

Turning to fairness, most participants believed that individual contributions to the tasks were roughly equal, which does indicate some initial tendency towards fairness in the work process. Yet most participants were not all that concerned about fairness—most were more worried about finding the required information and ensuring that their partner had some say in the outcome, as opposed to ensuring strict fairness through each person finding the same number of sources. These results emphasize two aspects of fairness in collaboration. First, the importance of context: our pairs had existing relationships which were very personal. Friends, for example, will have known each other for a reasonable period of time and may be more forgiving of one another's free-riding than they would be in professional work groups. Thus striving for fairness likely depends on the people with whom one is working and the circumstances in which work is completed. This latter issue was elaborated by a second aspect of our results, in that perceptions about the work itself were important. Several participants reported that the collaborative information seeking task didn't really feel like 'work', suggesting that these tasks were not perceived in the same light as professional, work-related duties. We suggest these issues had an impact on the extent to which participants needed to 'police' conformity to fairness norms.

A final aspect related to fairness was that, again, participants hinted at the difficulty of making accurate judgements about contributions. We have already discussed the fact that most did not make use of the shared search history due to interface issues. Participants resorted to making rough judgements about one another's work, due in part to the costs of information access at the interface. (We must remember that participants' perceptions of equality are based on retrospective self-report; it is possible that, if more had delved into the shared search histories, these perceptions would be less immediate.) Nevertheless, these results indicate the difficulty individuals encounter when trying to make judgements about the contributions of their collaborators. We regard the issue of judging contributions as an opportunity for design that raises myriad questions: How can we assist people in making fairness judgements when working through collaborative systems? What information needs to be presented? And what issues might occur when information is available to allow for fairness judgements? These questions are explored in the following chapter, where we examine granular awareness metrics in an existing collaborative work context.

6.6.3 Study Limitations & Future Work

This study is not without its limitations, though most of these can easily be regarded as opportunities for future work. An immediate limitation is that we were only able to report on a small number of pairs from a single culture. Despite an open call for participants, we did not receive any interest from groups of three or more. It is possible that larger groups would differ in terms of their search and sharing practices, calling for different features than those suggested here. Fairness might also be much more difficult to manage and gauge when more individuals are involved in the collaboration.

Our participants were mostly of university age and were relatively computer-savvy. Studies have shown that demographic and socio-economic factors can impact search behaviour and system usage (Aula & Haki, 2008; Hargittai, 2006) and these issues may be worth exploring in the context of collaborative search.

Regarding search tasks, the results of this study were based on tasks with high external validity, yet the patterns of behaviour reported here may not generalise to all other forms of collaborative search. For example, we were not able to examine the task of literature search and review, an endeavour that will no doubt be familiar to readers of this thesis. Several systems have been designed to support collaborative review, ranking, and re-finding of electronic documents (e.g. Capra *et al*., 2012; Golovchinsky *et al*., 2012). Future work should examine how our findings map to this commonplace and cognitively demanding task.

Studies have shown that the use of portable devices such as phones and tablets is an increasingly common method of conducting CIS (Morris, 2013). We did not account for this emerging aspect of collaborative search. Future work might examine the way in which searchers transition between use of different devices—one might imagine that searches on a mobile device would impose design challenges beyond those of desktop systems.

Finally, we were unable to access quantitative data collected at the system level—future work might employ such an approach to reach more precise measures of tool use. Reliable data about which features are, and are not, used in field settings would help to confirm our findings and guide the design of collaborative search tools.

6.7 Chapter Summary and Conclusion

This chapter reported a qualitative study of collaborative information seeking and tool use. We deployed two existing collaborative search tools (Coagmento and Diigo) to pairs of searchers with genuine, pre-existing collaborative information needs. Searchers then gave us their feedback on these tools and provided insights concerning the search process. This study provides insight into the process of real-world collaborative search, in terms of how search products are employed towards the process of shortlisting and arriving at an end product. Additionally, this study provides a number of pragmatic implications that can be used by designers to improve the interfaces of collaborative information seeking tools.

Additionally, this study has extended the earlier findings of this thesis as follows. First, our earlier work in Chapter 5 identified several strategies employed by searchers to manage redundancy. The present study confirmed the use of these strategies in real-world contexts, further emphasising their utility to designers of collaborative search systems. Second, we explored participants' approaches to division of labour and their concerns for fairness. In short, participants responses speak to the variable importance of fairness with respect to the task—searchers seemed less interested in ensuring equity, and this seemed to be related to the fact that they did not view the task as 'work' that would be completed in professional contexts. Thus their concerns for fairness were certainly more relaxed than those of the students in Chapter 3. Participants expressed some opinions that align with our prior findings in other ways, one of which was that it is very difficult to make judgements about a partner's work without sufficient information about their activities (cf. Galegher & Kraut, 1990; Kim *et al*., 2012). Thus, while searchers were not that concerned about fairness and did not feel a need to see the fine details about what their partner had done, the issue of judging each person's contributions to the shared task remains problematic. The following chapter draws on the findings of the studies reported thus far to consider how collaborative systems might support the monitoring of joint contributions, in turn allowing collaborators to judge whether each individual's contribution is fair.

CHAPTER 7

FAIRNESS THROUGH AWARENESS: CONTRIBUTIONS IN COLLABORATIVE GAMING

7.1 Chapter Overview

Prior chapters of this thesis have investigated the relationship between division of labour and fairness in the context of collaborative work. Literature reviewed and results obtained from Chapters 3–6 suggest that collaborative groups often desire, and consequently strive for, fairness in the division and completion of their work. However, achieving fair outcomes can be problematic for groups, especially in computer-mediated settings where monitoring joint progress may be inherently difficult due to limited awareness (Galegher & Kraut, 1990; Gutwin & Greenberg, 2002). Such concerns have also been raised in earlier chapters of this thesis; participants in Chapters 3 and 6 found it difficult to make judgements about fairness due to a lack of awareness about the ongoing contributions of team members. This chapter aims to draw these strands of work together in order to consider how collaborative systems could allow people to work in line with fairness preferences. We suggest that this could be achieved through awareness mechanisms designed to provide granular information about individual contributions.

We begin by reviewing the subject of awareness in collaborative work and problems associated with maintaining awareness in computer-mediated settings. We find that prior approaches, e.g. social translucence (Erickson & Kellogg, 2000; Erickson *et al.*, 2002), often aim to balance awareness with privacy during collaboration. However, we argue that such approaches would be insufficient for gauging fairness due to a lack of detail that does not allow for comparison between each individual's contributions. We suggest that, in order to assess fairness, collaborators might be provided with an awareness mechanism that provides detailed information about each team member's contributions. We investigate the potential for such awareness by studying an existing collaborative context where teams make use of information about contributions; in this case, the online game World of Warcraft (WoW). WoW is a large virtual environment in which players must collaborate in groups of 10–25 to overcome difficult challenges known as 'raids'. Raids are intrinsically collaborative, computer-mediated tasks that require a high degree of coordination, planning, and skill in execution. Fairness is also a concern for these groups; players must decide how much effort to invest during a given encounter, and also need to make decisions about the assignment of rewards from group activities. These aspects of raiding, combined with an interdependent division of labour among players, place WoW within the scope of CSCW (cf. Nardi & Harris, 2006; Bardzell *et al.*, 2008) and suggest a potentially productive context in which to study contribution awareness.

The study in this chapter considers how *damage and healing meters* contribute to the management of WoW raids. Such meters provide detailed visualisations of two key outputs for these groups. By studying how players make judgements about team member contributions, we aim to consider whether awareness mechanisms like those used in WoW would be applicable to other CSCW situations where fairness might be important. The contributions of this chapter are threefold. First, we introduce the idea of contribution awareness and identify WoW as a context in which similar information is already used. Second, by studying raiding in WoW, our study sheds light on the way in which awareness information contributes towards various aspects of raiding activity. We describe how damage and healing meters contribute to individual and group awareness, and also highlight sociotechnical issues stemming from the use of meters. Finally, we consider the extent to which these issues might apply if meters were adapted to other contexts. Thus, in aiming to distil our earlier findings into a simple design paradigm for collaborative systems, we contribute an understanding of how awareness information supports the practice of raiding in WoW.

7.2 Background

The following sections offer a short review of literature related to awareness in CSCW. This leads us to our consideration of how fairness judgements might be supported through contribution awareness. The concept of awareness was reviewed earlier in this thesis (see Chapter 2, subsection 2.3.3) and is generally related to knowledge about one's collaborators and their interactions with artefacts relevant to the collaborative task. Such knowledge provides a context for individual activities (Dourish & Bellotti, 1992), but the importance of the workspace itself is emphasised by Gutwin & Greenberg (2002), who state that awareness pertains to:

> "the up-to-the-moment understanding of another person's interaction with a shared workspace... [involving] knowledge about where others are working, what they are doing, and what they are going to do next." (Gutwin & Greenberg, 2002, p. 412).

Below we offer further insight into the utility of awareness for collaboration and identify why it can be difficult to support within CSCW systems.

7.2.1 The Importance of Awareness in Collaborative Work

Knowledge about what others are doing in a workspace supports many basic activities during collaboration. These include, and are certainly not limited to: assignment and coordination of work, management of coupling, communication about the task, anticipation of actions, and indentifying opportunities to provide mutual support (Dourish & Bellotti, 1992; Gutwin & Greenberg, 2002). More generally, awareness helps to "reduce effort, increase efficiency, and reduce errors for the activities of collaboration" (Gutwin & Greenberg, 2001, p. 9). These issues mean that awareness is essential for effective collaboration—without sufficient awareness, collaboration can be laboured and difficult to sustain (Gutwin & Greenberg, 2000).

Humans are very adept at gathering information necessary to build awareness, so much so that awareness is typically taken for granted in face-to-face settings. In ordinary circumstances, most people find it very easy to collect information required for answering basic questions about what is happening in a workspace: *who* is nearby, *what* they are doing, *where* they are working, and so forth (Gutwin & Greenberg, 2002). It is easy to collect information relevant to these issues when we work with others in the same space, and a variety of perceptual cues help us in this regard. Gutwin & Greenberg (2002) identify three primary means through which awareness information is gathered:

1. *Bodies and consequential communication.* The positioning of bodies in the workspace can provide relevant information; simply watching people work is enough to know where everyone is and what they are doing. Additionally, as people interact with artifacts and each other, *consequential communication* occurs (Segal, 1994). Such communication is not explicitly intended to provide information but does so as a consequence of action. An example offered by Segal (1994) is *visual monitoring*. When a pilot watches his partner manipulating a display, the very observation of action may be enough to know that certain procedures have been followed. In other words, information is obtained by watching the interaction, rather than by viewing the content of the display itself.

2. *Artefacts and feedthrough.* The use and movement of artefacts can provide awareness information. For example, objects may make sounds, change colour, or react to manipulation (Gaver, 1991). When an object is used, outside others can see or hear changes as a result of the interaction; thus, what would normally be feedback to the user also offers *feedthrough* to inform observers about actions (Dix *et al.*, 1993). The movement of objects among collaborators can also indicate task progress, ownership, or suggest an intended course of action (Pinelle *et al.*, 2003).

3. *Conversation, gesture, and intentional communication.* Verbal conversation is used to numerous ends during collaboration, from informing about progress through to clarifying problems and coordinating division of labour (Mark *et al.*, 1996). Conversation also allows for *overhearing*; people can 'listen in' on exchanges made by others (Heath & Luff, 1992) which may then be helpful in the completion of other tasks or in promoting peripheral participation (Monk & Watts, 2000). Qualities of speech like *intonation, emphasis, hesitancy, and pitch* can offer information about collaborators' feelings or state of mind (Gutwin & Greenberg, 2002). Information is also communicated through *verbal overshadowing*, which occurs when people speak their actions aloud (*"I am just placing this object here..."*) without an intended audience (Heath *et al.*, 1995; Roschelle & Teasley, 1995). Such utterances are useful because they unintentionally convey information to others nearby. Finally,

visual actions and *non-verbal communication* (gestures, gaze, facial expression) help to promote awareness: simple head nods can act as *emblems* to replace communicative acts, and gestures can emphasise or give clarity to spoken statements (Gutwin & Greenberg, 2002). Physical movements facilitate mechanisms of *deixis* (pointing and reference, e.g. *"that one over there"*) (Wong & Gutwin, 2010). The movement of hands in a shared task also supports coordination and is revealing about individual actions and intentions (Tang, 1991).

The immediate availability of the information described above, together with the inherent observability of actions, makes face-to-face collaboration a very natural and fluid experience (Gutwin & Greenberg, 2002; Moore *et al*., 2007a). When working in the same room as one's colleagues, it is usually quite easy to tell who is around and get a sense of what they are up to. Similarly, conversations will be straightforward, gestures will be readily visible, and the status of shared objects will be easy to ascertain. However, the information necessary for creating and maintaining awareness is not always available when collaborators are distributed, and this can make collaboration difficult to sustain. This is also true of many computer-mediated work settings, making the subject of awareness a critical problem of interest for CSCW.

7.2.2 CSCW and the Problem of Awareness

As noted above, awareness can be difficult to sustain outside of face-to-face interactions. This is especially true for computer-mediated work among remote collaborators who lack access to information that is often implicit for face-to-face teams (Herbsleb & Grinter, 1999; Gutwin *et al*., 2004). A problem specific to digital workspaces is that interaction within the shared space provides less information than equivalent actions in physical settings (Gutwin & Greenberg, 2002). For example, moving a PDF around one's digital desktop does not create the same shuffling noises (or *feedthrough,* Dix *et al*., 1993) that others would hear when shifting papers around a physical desk. Furthermore, the lack of presence information means that collaborators may find it difficult to establish basic understandings of who else is in the workspace, where they are working, what they are doing, and so on (Gutwin & Greenberg, 2002). In short, collaborative systems often limit "what people can perceive of others in the workspace" (Gutwin & Greenberg, 2002, p. 420) and reduce a person's "visual field to the limited area of a computer screen" (Gutwin *et al*., 1996a, p. 1).

The lack of awareness information means that distributed and computer-mediated teams have to invest additional, and often costly, effort to ensure smooth collaboration. Such efforts are overheads that go above and beyond what is required when working in co-located settings, and are problematic given that awareness is not typically a primary goal; rather, it is a secondary goal necessary for completing some task in the environment (Gutwin & Greenberg, 2001, 2002). Such concerns have made the the job of supporting awareness one of the most, if not *the* most, important research problems in CSCW. Yet designing effective solutions has proven to be an obdurate problem that remains a topic of interest (e.g. Kim *et al*., 2012; Kolfschoten *et al*., 2013). Because the related literature is large and varied, and a full review is not necessary for our purposes, we shall consider why supporting awareness can be difficult and why approaches have been hindered by problems related to privacy and excessive information sharing.

7.2.2.1 Supporting Awareness: Identifiability and the Privacy Tradeoff

Having discussed what awareness is, why it is important for collaboration, and why it can be difficult to acquire in computer-mediated settings, it is now worth considering some design solutions that have been proposed within the literature. Our aim here is not to discuss every single awareness mechanism ever created; rather, we seek to consider the issues that may arise when attempting to support fairness in the division and completion of collaborative contributions. Such issues could affect the acceptance of any system designed in this regard.[32]

As we have seen, awareness is critical for collaboration but can be hard to maintain in digital workspaces. This was especially true of early CSCW and groupware platforms, which had little to no support for awareness. Developing awareness mechanisms was technically difficult and for many years "people were aware of what collaborators sent them, and little else" (Grudin & Poltrock, 2012). The increase in computing power that arrived in the early nineties allowed researchers to begin exploring support for the sort of awareness that had, up until then, been almost non-existent (Grudin & Poltrock, 2012). Media space systems were an early example in this regard (e.g. Mantei *et al.*, 1991; Gaver, 1992; Bly *et al.*, 1993). Such systems included high-fidelity (by 1990s standards) audio and video links that attempted to recreate the co-located and informal nature of a shared office space. However, the perception of video as a solution to remote awareness retreated as it became clear that surreptitious observation of colleagues was not always desirable (Hudson & Smith, 1996; Grudin & Poltrock, 2012). A particular problem is that 'always-on' awareness feeds, as characterised by video links, blur the distinction between private and public channels (Hudson & Smith, 1996). People were seen to absent-mindedly engage in behaviours that are inappropriate for public space (Greenberg & Rounding, 2001), making hi-fidelity feeds especially problematic for use in spaces that are *ordinarily* private, e.g. the home (Greenberg & Rounding, 2001). In short, then, *too much* awareness information risks an invasion of privacy, especially when that information is being made public.

The issue of privacy makes designing awareness mechanisms a difficult task. One approach to managing privacy is *reciprocity*—making sure that whenever someone can see you, you can see them too. However, Hudson & Smith (1996) note that this approach has caveats in that it forces all spaces to become public and can cause excessive disruptions. They present some design approaches that obfuscate certain details in awareness feeds to allow observers to see *that* someone is there without necessarily being able to see *what* the person is doing. Other authors suggest high-level guidelines, in the vein of design paradigms, that attempt to balance awareness and privacy. One such approach, which we will use as a jumping-off point for our own considerations about contribution awareness, is *social translucence* (Erickson & Kellogg,

[32] For a review of how various systems have addressed specific elements of workspace awareness, see Gutwin & Greenberg (2002).

2000; Erickson *et al*., 2002). The goal of a socially translucent system is to make shared actions visible while supporting awareness and accountability. Regard for privacy is shown by the fact that the approach champions 'translucent' over 'transparent' design; a translucent system is one that displays who is around, and perhaps hints at what they are doing, without surrendering the fine-grained details that cause privacy problems. Such an approach necessarily entails making certain aspects of collaborators' behaviour visible while obfuscating others.

Figure 7.1 shows the *Babble* chat room system, a canonical example of socially translucent design (from Erickson *et al*., 2002, p. 41). Babble contains a buddy list (left side, Fig. 7.1) where the colleagues of the current user are listed using coloured dots. The centre box then contains the 'Commons Area', a social proxy which provides clues about the presence and activity of potential collaborators. This proxy acts as an indication of whether collaborators are available for conversation. When a person is active, his or her dot moves towards the centre of the circle. If a person is inactive, their dot resides at the edge of the area. And if a person is logged on but in another room (shown in the list on the right), their dot resides outside of the circle. Thus the system makes it possible for collaborators to be aware of *who* is around, *where* they are, and whether or not they are active, but does not reveal detailed information about each individual's current work activities.

Of course, social translucence represents just one point on the awareness–privacy continuum, and we do not suggest that is the *only* way of supporting awareness. Rather, we consider this approach because it helps to highlight the way in which judging equity and fairness might be difficult in a collaborative system. For example, a system like Babble might allow us to answer some basic questions about our collaborators, but the knowledge that others are 'active' is likely insufficient for judging fairness. Instead, we argue that collaborators require more detailed information about the contributions from each team member in order to make judgements about fairness. Such information might relate to *what* others are doing, *how* they are doing it, and perhaps also *how much* they are doing. The following section presents some considerations in this regard.

7.2.3 Towards Contribution Awareness: A Social Proxy for Effort?

Based on findings presented earlier in this thesis, we wish to consider how computer systems could enable collaborators to strive for fairness when completing divided work. We suggest that designers could achieve this by providing information about the *contributions* that are being made by each member of a collaborative team. In line with our earlier work and definition of fairness, we shall focus on the issue of allowing collaborators to make judgements about equality in contributions. Such a goal will intrinsically support fairness by allowing collaborators to observe deviations from equality; subjective judgements can then be made in line with individual expectations or whatever norms and procedures the group is following. While this may seem a relatively straightforward task, we believe that such a mechanism would need to account for certain subtleties concerning fairness and collaborative work in general. Here we aim to bring some specificity to the idea of contribution awareness and consider how it might be implemented.

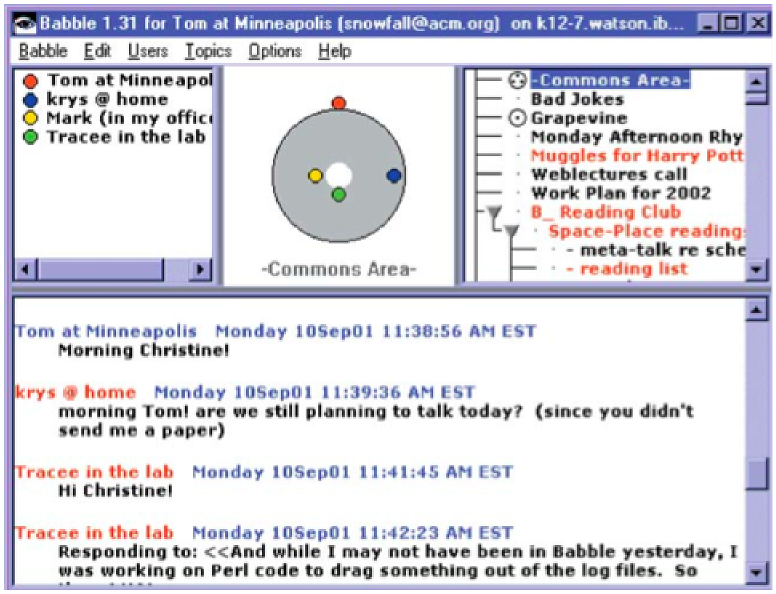

Figure 7.1: User interface of the Babble chat system, an example of socially translucent design, (from Erickson *et al.*, 2002, p. 41). The Commons Area (upper centre) is a social proxy that provides clues about the presence and activity of collaborators. When a participant is active, their individual dot moves to the innermost point of the circle. Collaborators unavailable for conversation are placed outside of the circle.

First, it seems intuitively obvious that collaborators would some need awareness of the very fact that contributions are being made. However, we contend that such general understandings must be accompanied by a reasonable level of *detail* about such contributions. This in turn would promote some understanding of actions, as opposed an understanding of mere ‘activity’. For instance, the message that ‘Ryan is typing’ offers a hint about the author’s current work task. But the message that ‘Ryan is typing at 70 words per minute in Chapter 7’ makes it easier to specify current performance. One way to think about this is in terms of a continuum between *ambiguity* and *precision*; more or less information can be provided about a person’s work that, in turn, can help facilitate judgements about equality. If sufficient detail about work is available, one might be able to answer questons about whether particular levels of performance are appropriate to the given situation. We cannot prescribe an exact level of precision in this regard; this would very much depend on the task at hand, and the challenge for the designer would be to determine the correct level of detail at which information is sufficient to allow for equity to be realised.

Our second contention is that an awareness mechanism for judging equity (and hence, fairness) will need to allow for *social comparisons*. As we have seen, fairness is intrinsically social and people make judgements about equity by assessing their own outcomes relative to those of others (cf. Festinger, 1954; Knez & Camerer, 1995; Fehr & Schmidt, 1999; Cohn *et al.*, 2009). Thus, judging equity in a collaborative work setting would require that people are able to see their contributions vis-à-vis those of others. The most straightforward way of achieving this would be to provide some shared representation of contributions, with each individual’s contribution shown relative to those of everyone else. Such a representation would need to be *reciprocal* in that collaborators would have to share details about their own efforts in order to receive information about those of others. Again, we cannot specify the correct content for any such representation, and nor can we prescribe the exact means by which relative fairness should be assessed. Such issues are very much dependent on the types of contribution that are relevant to the task at hand, and it might be necessary to alter the focus of such a display based on critical aspects of the group task (cf. Straus, 1999). In brainstorming, for example, the number of ideas contributed by each person might be a suitable metric, whereas in tasks with well-defined stopping rules, measures of individual progress or intensity of work (e.g. words per minute) could be more appropriate.

It is important to note that we see these contentions as tentative and by no means definitive. We do not suggest that mechanisms for judging fairness in contributions should be the *only* source of awareness information available to a team. Rather, we suggest that mechanisms for assessing contributions should extend whatever information is available for answering the basic questions about awareness: *who* is around, *where* they are, *what* they are doing, and so on. We envisage a mechanism in the style of an awareness widget—an interface element that augments a digital workspace with information about the actions of collaborators (Gutwin *et al.*, 1996a,b; Gutwin & Greenberg, 1997). Very much in line with the general scheme of this thesis, we do not suppose to tell people what is fair or what they should do. We are simply concerned with exploring the challenge of providing the necessary information that could allow fairness judgements to be made. Nor do we seek to argue that *all* collaborative systems should support fairness—this is a matter that should be decided in accordance with the situation at hand.

The ideas presented here inevitably raise some questions about the way in which awareness of contributions might impact a group's work. The fact that computers can keep a persistent record of contributions might actually make fairness easier to manage than in offline settings, especially if each individual can be held accountable for his or her actions. However, requiring people to surrender information about their contributions, and possibly their work process, necessarily infringes on their privacy. Whether or not this erosion would be problematic in a group setting is another question. Several prior studies do suggest that people are sensitive to increases in information disclosure and are especially wary of unwarranted personal monitoring from others (Ramage, 1994; Harper, 1996; Cheverst *et al.*, 2007). Yet increased information disclosure might be felt less acutely if the information is clearly task-related, is provided with the consent of all team members, and leads to collective benefits that are recognisable to all team members. It would also be interesting to examine the impacts contribution awareness might have on group dynamics and overall participation. For example, a study by Harper *et al.* (2007) found that contributions to an online movie ratings site increased by 530% when information about the median contribution was emailed to community members. However, those above the median actually lowered their overall contribution, implying that providing a target for performance had both positive and negative consequences for collective participation.

While our present considerations have aimed to abstract away from specifics about how contribution awareness might be implemented, we note that there are some existing circumstances in which mechanisms like those we have proposed are already in use. One such setting, described in brief at the outset of this chapter, is in the online game World of Warcraft. In the remainder of this chapter, we report a study of groups in Warcraft. Our study explores how groups use simple *meters* that quantify and visualise *damage* and *healing*, two primary outputs integral to task progress. We were interested in how meters contribute to different aspects of group awareness, in terms of making individuals accountable for contributions and in ensuring groups meet their targets. In this way, the study is akin to an evaluation of our design idea—we aim to understand how meters can be beneficial, but also explore the various sociotechnical issues that arise through their use. The following sections offer some background on grouping in Warcraft and lead into our consideration of damage and healing meters as mechanisms for contribution awareness.

7.3 Thesis Study 7: Awareness Meters in World of Warcraft Raid Groups

7.3.1 Warcraft Raiding: Background

World of Warcraft (hereafter, WoW) is a massively-multiplayer online game (MMOG) with approximately 9 million active users around the globe (Birnbaum, 2013). In WoW, players choose a character and enter into a virtual environment populated by various fantasy races (elves, orcs, gnomes and the like). The goal of the game is to complete quests and battle enemies to obtain experience points. After a while, experience points result in a ‘level up’ that increases the character’s strength, in turn furthering the player’s ability to progress through the various areas of the game world.

Like other MMOGs, WoW emphasizes the collaborative and social aspects of gaming. This is especially true at higher levels, where players must band together in teams of 10 or 25 to defeat challenging dungeons known as “raid” encounters (Bardzell *et al*., 2008, 2012). The basic challenge of a raid involves progression through a series of battles with increasingly tough ‘boss’ monsters (Bardzell *et al*., 2012). Defeating a given boss allows players to acquire powerful rewards (known colloquially as ‘epic loots’) which enable further progression by, for example, bestowing players with the strength required to face more difficult bosses and dungeons. Raids also offer players the opportunity to accumulate social capital such as status, reputation, and friendships (Jang, 2007). This capital then supports further progression while fostering an experience of fun and achievement within the context of the game (Jang, 2007; Bardzell *et al*., 2012)

However, navigating a raid dungeon is by no means straightforward. Not only are raids time and labour intensive, the majority are punishingly difficult and represent the uppermost challenges within WoW.[33] Most raids require a high degree of planning, coordination, and communication among players, and groups may need to adhere to a specific strategy (or set of strategies) to defeat a given boss. Additionally, players often need to manage transitions between different phases of an encounter, with each phase requiring a new tactic that players must adopt in order to progress. Failure to perform these actions can place the group in jeopardy, sometimes resulting in a “raid wipe” that leads to death for the entire group

[33] As the reader may not be familiar with the finer points of WoW raiding, Appendix I provides an example of a raid’s progression through an encounter with the dragon *Onyxia*.

and requires the encounter to be restarted. Group members must therefore ensure that their individual contributions are consistently well-timed and appropriate to the required strategy. Moreover, because the line between success and failure is very thin, raids cannot afford to tolerate non-participation or free-riding behaviour. Both of these issues entail a certain degree of awareness between raid members—team members must monitor individual and collective contributions, making raiding highly relevant to our interest in contribution awareness. Although there has been some prior work on awareness in WoW and other virtual environments (e.g. Brown & Bell, 2004; Ducheneaut *et al.*, 2006; Moore *et al.*, 2007a,b) the question of how groups monitor collective contributions during high-intensity raids has not been explored. Thus, in our exploration of damage and healing meters, we are presented with the opportunity to make a contribution to the literature on WoW and collaborative gaming in general.

A second aspect of raiding relevant to the current work concerns the assignment of rewards. As mentioned above, raids are undertaken in pursuit of loot drops, but the distribution of loot from raids is almost always uneven—only a limited number of items drop from each boss, and not all items that drop can be used by all players. This means that no single player is guaranteed to be rewarded after each encounter. Groups are therefore faced with the problem of deciding how to best allocate rewards while accounting for the contributions of all raid members. Such decisions may be taken in accordance with fairness norms. Studying the use of awareness metrics in WoW offers an opportunity to shed light on how work is divided in an additional, non-rarefied setting.

7.3.1.1 Division of Labour in Raiding

As a collaborative endeavour, raids necessarily entail division of labour among their members. This division of labour is managed through *ludic* (meaning 'game defined' Bardzell *et al.*, 2012) and *social* roles. Regarding the latter, most raids have a leader who acts as a kind of project manager: he or she is responsible for assigning sub-roles before the start of the raid, outlining attack strategies, and keeping track of overall progress. Other social roles include the communications officer, responsible for issuing instructions during battles, and the loot master, a player who takes temporary possession of the items a group receives from each boss. Ludic roles, on the other hand, are defined by each player's character class. Different classes have different abilities, and these in turn define a player's responsibilities within a raid group. Some classes, like Warriors and Death Knights, are adept at mitigating damage and protecting other players from enemies.[34] Priests and Shamans are primarily responsible for healing other players and providing benefactions. Classes like Warlocks, Mages, Hunters, and Rogues focus on offense, applying 'damage-per-second' (DPS) required to defeat the raid boss. Finally, some other classes blur the lines between these three roles: Paladins and Druids, for example, can tank, heal, *and* apply damage, though not all at the same time—players typically need to focus their efforts such that they specialise in a single one of these roles, rather than all three.

While many encounters call for a different division of labour among players, the interdependencies

[34] Holding the attention of an enemy in this way is known as 'tanking'.

among ludic roles define the division at a very high level. Characters responsible for tanking monsters will not live long without the aid of a healer, and neither healers nor tanks can defeat a boss without the support of DPS classes. This means that, during group activities, particular classes are accountable for certain tasks but not others. For instance, it is accepted that the job of the healers is to keep everyone else alive; thus, if other characters suffer death during an encounter, suspicion of fault is immediately directed towards the healers (Bardzell *et al.*, 2008).

7.3.1.2 Awareness in Raiding

Although there is an emerging body of work on WoW and virtual environments more generally, no academic work has yet addressed the subject of awareness in raiding. This is surprising given that, as an intrinsically collaborative task, raiding must entail a certain degree of awareness among players. Many features of raid battles are indicative of this need. For example, some encounters require players to ensure that a particular spacing is maintained, else the raid may take additional damage from certain triggered effects or threats in the local environment. Performance of such coordination must require awareness of others. Additionally, players may have an understanding about what each raid member's role is, but will nevertheless require an up-to-the-moment understanding of whether or not specific tasks are being performed. We identify three methods of obtaining awareness information in WoW:

- First, explicit communication is used to create and maintain awareness. All players have access to a series of text-based chat channels within WoW, at least one of which is private to the raid group. Because text-chat is difficult to use when raiding—players need to use their hands to control their characters— the majority of raid groups use out-of-band voice chat systems, e.g. Skype, TeamSpeak, or Ventrilo, to allow for voice communications (Chen, 2008).

- Second, as a mimetic representation of its user, an avatar naturally gives off information useful for awareness as it moves around the virtual environment (Gutwin & Greenberg, 2002). For example, a player's position in virtual space may be revealing about current or intended courses of action. Avatars often animate while executing certain actions, e.g. spellcasts involve different movement of the avatar's limbs, and these animations can be recognised by other players. Avatars are sometimes expressive enough to communicate emotions; for example, typing 'LOL' in Warcraft causes the player's avatar to tilt its head and laugh out loud. At a higher level, the fact that collaborators are represented by avatars means that an individual need only survey their immediate field of view to answer basic questions like who is nearby and whether or not they are enemies.

- Third, WoW has a number of interface elements that support basic aspects of awareness, e.g. a minimap that shows the player's current location relative to other players in the environment. Mousing over this map reveals the identities of said players and indicates whether each is friendly or hostile.

Prior work on awareness in virtual worlds has identified some deficiencies with the above methods (Brown & Bell, 2004; Moore *et al.*, 2007a,b) but one issue that has not been covered is the fact that these channels are usually insufficient for raids. This is partly due to their limited bandwidth—for instance, in a group of 25, not all players can speak aloud at the same time, and text chat is easily flooded with so many participants—but also because they do not provide all of the information required by raids. To acquire such information, players use a variety of third-party interface 'add-ons' to augment the basic interface of WoW. Many such add-ons exist to support raiding, and here we shall consider *meters* that display information about the damage and healing performed by members of a raid group. Use of such meters appears to be prevalent within the wider WoW raiding community, and we will consider how these meters support awareness of contributions.

7.3.2 Damage and Healing Meters: Simple Metrics for Contribution Awareness

Figure 7.2 displays two examples of damage meters from WoW. We can see that these meters have several properties that align with our earlier contentions about contribution awareness. The first is that meters present an aggregation of the damage or healing performed by each player in the group. This provides a rough idea of each player's overall contribution to the collective effort. The second is the hierarchical ordering of players using individual bars that allow for *social comparison*, i.e. assessment of one person's contribution relative to the next. Players are further distinguished by the colour of their individual bars. Colour in this case pertains to each person's class—green is typically associated with Hunters in WoW, Rogues appear as yellow, etc. These colours are consistent with other elements in WoW and are thus familiar to the majority of players.

Additionally, we see that meters can be progressively more or less *detailed* in terms of the information they present. Meter A in Fig. 7.2 shows a simple bar chart with each raid member's name; this information is slightly inexact but nevertheless offers a rough idea about each player's contribution. Meter B makes this information more explicit by providing exact figures that show (Fig. 7.2 B, left to right) the total damage inflicted by each individual; their damage per second; and their contribution as a percentage of the overall group effort. In this way, the second meter allows players to not only compare their contributions but also quantify the difference between rankings.

While Fig. 7.2 displays only damage meters, healing meters are functionally and aesthetically identical to those used for damage. The only difference is the type of information that is presented by the meter. However, most healing meters provide an alternative view that allows players to view *overhealing*, which captures the amount of healing that was performed unnecessarily on players already at full health. While a small amount of overhealing is unavoidable, excessive overhealing can be regarded as a sign of poor play. Rather than making calculated decisions about where to direct effort, a player may be healing the wrong targets or 'spamming' the same spell over and over. Either of these behaviours would exaggerate his or her contribution on the basic healing meter. Overhealing is also distinct from damage dealing because it is not possible to 'overdamage' an enemy, making overhealing an important source of information about the

Figure 7.2: Two damage meter addons for World of Warcraft. (A) shows a basic damage meter. (B) shows the *Recount* interface addon, which provides more granular information about the total damage inflicted, damage per second, and percentage contribution to the group's overall total.

quality of work that is being done.

In discussing these meters, it is important to recognise that the information they present is not normally available to raid groups. In ordinary circumstances, players know only their own efforts and have no way of finding out how much damage or healing their comrades are contributing. Meters work by extracting this information from each player's WoW client and presenting it in a shared visualisation that is made available to all members of the raid group (provided that they too have the relevant add-on installed). Thus, without meters, players ordinarily have no way of evaluating the contributions provided by each raid member.

We contend that meters embody the primary design properties that might be included in a mechanism for contribution awareness. What we wish to understand in the present study is how meters are used to support awareness in the context of raid encounters. It seems immediately obvious that the presence of damage and healing meters will alleviate the need to communicate explicitly about these issues, but the precise utility of these meters, especially for making fine-grained inferences about fairness, is currently unclear. The present study addresses these questions by interviewing players of WoW about meters and the issues that arise through their use. This allows us to unveil a range of sociotechnical concerns, which we then aim to extrapolate to the broader issue of awareness in computer-mediated workspaces.

7.3.3 Are Games 'Work'?

Given that most people play games like WoW in the pursuit of leisure, the reader may harbour reservations as to whether any insights obtained would be applicable to the broader context of other CSCW situations. Here we sketch some brief justifications as to why WoW is an appropriate setting for our study.

The first thing to note here is that games, particularly Warcraft, have been a topic of concern for the CSCW community within recent years—several published works provide a precedent to suggest that productive implications can be derived from the study of gaming contexts (e.g. Nardi & Harris, 2006; Bardzell *et al.*, 2008, 2012). Second, virtual platforms like WoW are not always perceived as games *per se*. The arduous and often repetitive nature of in-game tasks frequently blurs the line between work and play (Yee, 2006c; Chen, 2008). Games like WoW usually include professions (e.g. blacksmithing) which, if mastered, allow players to produce commodities that can be sold or traded; thus many players are unwittingly 'working for fun' (Yee, 2006c). Additionally, one survey of more than 5000 players (Yee, 2006b) revealed that 60% of respondents regularly spent 10 consecutive hours in an online world, with some spending 16 hours or more per day. This made the virtual world akin to a full-time job for these players, many of whom regarded the virtual world as both their job and their 'home' (Yee, 2006a).

At a broader level, the problem of distinguishing between work and play remains a topic of debate within the CSCW community (for examples, see Crabtree *et al.*, 2005; Schmidt, 2010). In line with our above statements concerning virtual worlds, one can think about play as a form of work, and of work as involving some degree of play (or at least, leisure, Brown & Baarkhuus, 2007). Reflecting this line of thought, the concept of 'productive play' has recently entered the CSCW lexicon. The argument that games are somehow irrelevant to CSCW is, therefore, problematic and rests on the notion that work and play are mutually exclusive—if one considers the concept of work as 'any purposeful activity, whatever the goal' (Brown & Baarkhuus, 2007), this would easily encapsulate gaming activities. Furthermore, it can be argued that although WoW does not result in a 'work product' beyond the intangible goods and experiences of players within the game, there is a collective output of *having fun* (cf. Nardi & Harris, 2006) and the fact that this is both collaborative and computer-mediated is enough to place WoW within the scope of CSCW.

Given these considerations, we are confident that findings derived from WoW can be extrapolated to other settings. The following sections describe the methods employed during this study and the main findings.

7.4 Method

To reiterate, this study aims to investigate how awareness meters in WoW support team participation and judgements about fairness in contributions. Additionally, we aim to understand the various sociotechnical issues that arise through use of these meters, with a view to considering how such issues might impact the application of meters in other collaborative systems.

Following the precedent set by earlier work on WoW (e.g. Bardzell *et al.*, 2012; Chen, 2008; Ducheneaut *et al.*, 2006), this study was informed on several levels. First, the author of this thesis is a former WoW player with first-hand experience of raiding. This participation frames our analysis and interpretation while providing an understanding of the social situation in which participants' collaboration occurs. Such an understanding is critical for game-related research, where "researchers should have played the games they are studying... if they do not, they cannot know what questions to ask, decipher the local language, understand the game mechanics, or have any sense of the social context of play" (Williams *et al.*, 2006, p. 342).

Second, we engaged in observation of players completing current high-level raid content in WoW. This was done to update the author's understanding of WoW and to improve our ability to question current players—several years have elapsed since the author's last period of protracted play. Observations were conducted by watching live, player-created web streams. These are essentially public video and audio feeds where players 'invite' others to watch them raid by broadcasting the contents of their screen over the Internet. Four such feeds were observed over four separate one- to three-hour sessions. (The exact feeds viewed are listed in Appendix H). We did not obtain consent to observe these players because they invite observation by making their feeds public. Additionally, we did not collect data about these players; rather, the author took notes about aspects of the raid encounter, e.g. what the coordination requirements were, what the name of the boss was, and so on. The researcher also noted the actions performed by players and the types of interface elements used. These notes, combined with the author's own experiences of WoW, helped shape the questions used in our third and final stage, where we performed semi-structured interviews with current WoW players. These interviews aimed to unveil players' experiences, and issues encountered, during use of damage and healing meters during WoW raids. The findings we report here are derived from this latter data source.

7.4.1 Participants

We interviewed two male and three female WoW players.[35] Interviewees' ages ranged from 18–46 (M = 26.8). Interviewees had been playing WoW for an average of 6 years (range = 4–8.75 years). All had experience of raid encounters ranging from early content in WoW's initial release through to its most recent expansion (at the time of writing, *Mists of Pandaria*). Perspectives from all currently available classes were represented in our responses, and our interviewees encapsulated tanks, healers, and damage dealers. Four of the five interviewees had experience with multiple classes and raid roles. Two players had also held formal, organisational roles within their raid groups, e.g. raid leader or communications officer (responsible for giving commands during raids).

All interviewees had been in a raiding guild. A guild is a persistent collective of anywhere between two to several hundred players (Williams *et al.*, 2006). Guilds are often synonymous with raid groups because raids are often comprised of players from the same guild (Chen, 2008)—effectively, the formal structure facilitates persistent grouping (Bardzell *et al.*, 2012) and also allows certain policies to be implemented that help to manage raids in the longer term, e.g. a guild bank to which all members can contribute. Two

[35] We recognise that five interviewees is a vey low number considering that WoW has more than 10 million active players, and thus our responses can not be considered as truly representative. However, one benefit of the low number of responses is that we were able to consider each person's response in depth, and we were beginning to see saturation in terms of topics covered even with just five responses. Since our aim is to illustrate salient issues associated with contribution meters, we feel that the present study has proven successful in this regard, even with the low number of interviewees.

of our interviewees described their guild as casual (one or two raids per week), two stated medium-core (two to three raids *every* week, without fail), and one as hardcore (three or more raids per week, without fail). Two had held positions of authority within their guilds. Additionally, all respondents had engaged in raids with their guild and in ad-hoc pick-up groups (PUGs). PUGs are one-off collectives that form to defeat an enemy and then disband afterwards. Even though our number of interviewees is low, these demographic aspects do suggest that the raid experiences surveyed were broad and not isolated to just casual or hardcore players.

7.4.2 Procedure

Interviewees were recruited through Facebook. Interviews were conducted one-to-one between individual participants and the author of this thesis. Two interviews were face-to-face and three were via telephone/Skype. All interviews lasted between 50 minutes and 1.5 hours. Interview questions used in this study can be found in Appendix H. We began each interview by requesting the demographic information described above. We then asked questions that covered a range of raid-related behaviours. These included how the interviewees determined appropriate actions in the context of a raid; the interface elements used to remain aware of teammates' behaviour; and how the interviewees determined whether others were contributing properly on raids. We also asked whether they used damage/healing meters (which all respondents did) and probed their opinions and experiences with these particular mechanisms. Interviews ended with a short debrief about the purpose of the study. Participants were not paid for their participation.

7.4.3 Analysis

All interviews were recorded and transcribed by the author of this thesis, allowing for early familiarisation with the dataset and development of initial understanding. This resulted in 52 pages of transcript. We used open, axial, and selective coding (Strauss & Corbin, 1998). The coding process was iterative; transcripts were given two complete readings and initial codes were evaluated and refined during each reading, with internal consistency strengthened by scrutinising the data for counter-examples. Codified concepts were then structured into the themes that comprise our results set.

7.5 Results

Our analysis produced eight clear themes, and these are described in the subsections below. For ease of interpretation, we separate themes into two distinct groups. Our first group considers themes relevant to *how contribution meters support raid groups*, ranging from a general sense of awareness through to managing fairness during distribution of loot after an encounter. Our second subsection then considers *sociotechnical issues* that arise through the use and interpretation of contribution meters. These themes cover both positive and negative issues.

Before presenting the themes, we first consider the extent to which awareness is necessary during raids, as evidenced by player responses that imply the general importance of awareness in WoW. Like our earlier qualitative analyses, we use quotations that are illustrative of particular points. We identify speakers by their participant number, gender, and age (M/F).

7.5.1 The Importance of Awareness in Raids

Participants' responses offer some insight as to why awareness is both necessary and important in raid settings. All participants considered it necessary to maintain a high-level sense of awareness of what is happening at any given time during a raid. One participant referred to this as 'raid awareness', which is perhaps best likened to workspace awareness (Gutwin & Greenberg, 2000, 2002) in that it relates to a very general understanding of the group's status and their interaction with various entities in the virtual environment. As the interviewee surmised:

> *"It's important to be aware of everyone around you—this includes healers, tanks, and damage dealers—so you can spot someone who may be standing in the wrong place, or may just be in a bad place in general...you need to focus your eyes on the centre of the screen so you're able to watch the boss and the ground for incoming attacks and keep an eye out for bad positioning. This gives you the ability to tell them to move, which will avoid wipes."* [P3, F, 19]

Raid awareness appears to underpin many different aspects of raiding. For instance, there is the need to be vigilant of, and subsequently control, unpredictable threats. Many encounters involve additional monsters, known as 'adds', that are weaker than the main enemy but are sufficiently dangerous that the raid must deal with them or risk a wipe. A second issue concerns the placement of team members within the virtual space. Many encounters inflict additional damage if raid members clump together or stand in particular areas; thus, the team needs to be aware of these issues and re-coordinate physical placement when necessary. Finally, there is a need to monitor progress on the primary task; certain encounters must be completed within a specific time frame, or else the enemy will 'enrage', resulting in increased damage that the group may not be able to withstand. Team members must therefore maintain a degree of overall awareness if they want to maximise their chances of success on the raid and minimise the threat of wipes caused by poor coordination or sloppy play.

Additionally, and in line with the earlier definition of awareness offered by Dourish & Bellotti (1992), raid awareness provides a context for individual actions, in the sense that observing the behaviour of others appears to enable raiders to make decisions about appropriate actions. This was elaborated by the same participant as above:

> *"Being aware of what everyone else is doing allows me to see when I need to use my own abilities, whether defensive or offensive, or to swap to a different position. If I see the rest of my raid move to the other side of the room, I need to move too. But if I wasn't raid aware I wouldn't be paying attention to them, and thus I'd end up probably dying from not moving."* [P3, F, 19]

Thus raid awareness is not only related to the interdependent division of labour among players but also allows players to adapt and react strategically in the high-pressure raid environment.

We now delve into the ways in which meters contribute to various aspects of this overall raid awareness. We also consider the extent to which meters are involved in decisions about fairness.

7.5.2 How Meters Support Raid Groups

7.5.2.1 Monitoring Contributions and Identifying Underperformers

As discussed earlier, the information presented in damage (or 'DPS')[36] and healing meters is not usually available to players when using only the standard WoW interface. Meters work by collecting information hidden at the system layer and forming an on-screen visualisation of this information. Metered content is then available to all members of the raid, provided that they have the relevant interface add-on installed.

The first, and perhaps most obvious, way in which meters support raiding is by allowing each member of the raid to actively monitor the contributions of all other individuals in real time. The presence of an individual on the DPS meter is often enough to know that they are active and contributing, in turn lessening the need for explicit communication about individual actions. Conversely, the *absence* of an individual can prove informative, especially if that player suddenly disappears off the meter—players can infer that the individual has some problem, e.g. loss of internet connection, or is distracted by another task, which may or may not be raid-relevant. Meters thus provide a very quick resource towards which players can glance to get an overall sense of the raid's status. Yet meters also allow for monitoring of individual participation. As one player described:

> *"Everyone must pull their weight to progress through raiding... In WoW all it takes is for one person to mess up and it's a wipe, so everyone must do the tactics and conform to their role to the best they can... DPS meters give raid leaders the information to see who is slacking or failing."* [P1, 28, M]

The latter sentence within P1's statement leads us to a second way in which meters are used; that is, to identify individuals who are underperforming. Recall that meters are ranked hierarchically according to individual outputs. Thus, those at the bottom are always contributing less damage or healing than those at the top of the meter. However, meters also allow players to compare contributions by gauging the relative size of each individual's contribution bar (see Fig. 7.2). Thus players can make judgements based not only on ranking but also on the basis of a quantified contribution measure. These dual aspects of meters help the raid deal with underperformers in various ways. First, the consistent placement of an individual towards the bottom of the meters can be used to identify opportunities for improvement:

> *"If you have four people all doing roughly around the same damage... and then two straggling really badly at the bottom you know it's to do with them, it's not to do with the raid or the boss."* [P2, 23, F]

[36] DPS is an acronym for damage-per-second, a jargon phrase that refers to 'dealing damage'.

In this way, meters allow raids to check for non-contributors, who may then be removed from the group based on evidence available in the meters. However, meters are not always enough to know *why* an individual is underperforming. This is important because a player's consistent placement near the bottom of the meter does not mean they are lazy *per se* but could instead mean that the player is in need of guidance. Interviewees described how the information presented in meters helped to identify such opportunities for improvement. Such improvement can be assumed to support the long-term health of the group, with those towards the lower end of the rankings encouraged to make improvements so that their contributions come into line with other raid members:

> *"In the raid it is particularly so you can identify weaknesses... it might be that someone's having a real problem with what loot they've got, what armour... so the aim isn't to say, 'oh get out of here we'll get someone new', it's to identify weaknesses and learn from them... so you can eliminate all the weaknesses."* [P3, 19, F]

> *"In rare cases it's down to using the wrong rotation*[37] *so we'd try to correct that and get them to look further into their class and learn the right rotation."* [P4, 46, F]

In general, the very presence of an individual on the DPS meters implies that he or she is actively participating in the raid. Interestingly, presence on the DPS meters cuts both ways. Sometimes, the fact that someone is on the DPS meter can be a sign that they are actually *not* performing the correct tasks. Many fights require certain players to control additional enemies using traps or immobilising spells, entailing a large drop in DPS contribution because the individual in question must divert his or her offensive capabilities away from the primary target. These 'crowd control' tasks tend to be pre-assigned, and thus if the person in charge of such a task remains present on the meters while additional enemies are around, raid members will be aware that the person in question is not performing the correct task:

> *"In this case if you weren't doing your job then you were high on the meter. It's how we knew the hunter wasn't doing his job, when he started to shoot up the DPS meters because he was just standing still and firing on the boss, instead of killing the add."* [P4, 46, F]

In this way meters allows players to be held accountable for both appropriate *and* inappropriate behaviours.

7.5.2.2 Providing Targets for Performance

A second way meters support raid activity lies in providing a target for performance. Because the information meters display is detailed, players are able to ascertain whether certain performance standards are being met. This can be of critical importance during certain scenarios where players must achieve, and subsequently maintain, a certain numerical threshold of performance throughout the encounter, else the

[37] A player's 'rotation' refers to the order in which his or her offensive abilities are executed. Certain rotations produce better damage output than others, making rotation selection an important aspect of raid behaviour.

raid may stand to fail. The damage meter provides a common resource for gauging whether this threshold is being met and who, if anyone, needs to increase their effort during the encounter:

> *"I often look at my damage meter to make sure I'm achieving the right amount of DPS, this applies to every raider so that we can kill the boss quickly. If the right amount of DPS is not met, then the boss can enrage after a certain amount of time, causing a wipe."* [P3, 19, F]

In some raid encounters, it is often unclear what the required level of performance for a damage or healing class should be. Studies suggest that people have a tendency to look for normative standards when behavioural requirements are uncertain (Gibbons & Buunk, 1999). In line with arguments earlier in this thesis, one way of acquiring such a standard is through social comparison—people look to the behaviour of others in order to guide their own actions, in essence 'doing as others do' (cf. Festinger, 1954). In WoW, meters are a resource for this information. While contributions on some raids are reactionary, in the sense that they are defined by what is going on in the raid and the abilities of the boss, the visualisation of performance gives a target standard to which team members can direct their efforts.

7.5.2.3 Facilitating Public and Private Comparisons

Responses suggest that a certain degree of equality in contributions is desired in raid groups. In line with our contentions earlier in this chapter, meters help players achieve equality through the visualisation of contributions relative to one another. As in Fig. 7.2, the position of each player's coloured contribution bar makes assessment of equality relatively straightforward:

> *"You could see that by a bar on the DPS meter, as long as bars are all aligned then everyone's doing roughly the same, of course there's variations but I don't really mind... it shows everyone's putting in a fair amount of effort."* [P1, 28, M]

Equality in contributions appears to be important for two reasons. The first is very much related to fairness and the expectation that all members contribute: if contributions appear similar on the meters, players can be assumed to be fulfilling their role and group leaders need not be concerned about monitoring individual performances. However, a second reason pertains to the overall risks of raiding, in that it is better to have even performance between members rather than a 'spikier' or more polarised output, where a few individuals pull away at the top and all others remain at the bottom. This is because the latter scenario is much riskier in a raid setting; if the top contributors disconnect or suffer death, the raid is at a higher risk of failure due to a larger overall percentage loss of damage output. Concerns such as these may explain why raid groups invest much effort in training new or underperforming players: it is in the collective interest for contributions to be fairly equal and of a certain (usually high) standard. Meters provide raid groups with the information required to assess both of these concerns.

7.5.2.4 Supporting the Assignment of Rewards

Recall that assignment of looted rewards is a problem faced by raid groups: items are limited in quantity and not all items can be used by every member of the raid. Moreover, two or more players may feel that they have equal entitlement to a given drop. These issues mean that decisions must be made about how to assign rewards from the collective effort.

Responses revealed a range of loot assignment policies within the guilds of players. Two players mentioned that their guilds used structured reward systems to assign loot after boss kills (cf. Chen, 2008). These systems reward players with a point-based currency that can be used to 'bid' on loot drops. The highest bidder receives the reward in question. One participant stated that his guild used meters to resolve tied bids:

> *"If there were two people who bidded the same amount of* [points], *then it'd be decided by the raid leader who would use their information through the DPS/Healing meters."* [P5, 18, M]

A different approach was to consider the item that had been dropped and balance rewards with overall benefits to the raid group. For instance, a particular item may be a minor upgrade for a regular player but could be a major upgrade for a weaker member of the team. Meters also factored into such decisions:

> *"In one of my guilds we use what we call a 'loot council'...officers in the guild take into account people's current gear and their damage and healing on the meter. If it's a big upgrade for the new player they will award it to them."* [P3, 19, F]

It is, therefore, not the case that being 'top of the meters', which indicates skilled performance and a significant contribution to the team, necessarily equates to a greater chance of receiving items. Rather, meters are used as evidence of specific needs or to justify potentially controversial loot assignments.

7.5.3 Sociotechnical Issues

In addition to the generally positive aspects of meters outlined above, we identified various sociotechnical concerns which, we believe, could have implications for implementing contribution meters in other settings.

7.5.3.1 Promoting Competition over Coordination

Being top of the damage or healing meter is often considered an accolade carrying a certain degree of prestige. Topping the meters implies that the individual in question has not only made the greatest contribution but is skilled to the extent that it lifts him or her above the rest of the group. Players appear to vie for the attention that this status brings, and one interviewee described how the 'top of the meters' accolade became a motivating target for members of her guild, creating a level of competition that was beneficial for progression:

> *"Someone's belting away a load of danage and there might be some friendly egging on in raid chat, like, oh look at me, see who can beat me. And then everyone else tries harder to beat that."* [P2, 23, F]

However, this same participant described how competition could also be problematic in that some individuals can become overly focused on achieving ‘top of the meters’ status. Two other interviewees also mentioned this side effect. Clamoring to be top of the meters can lead players to neglect the basic coordination duties demanded by raiding. Ignoring the optimal strategy just to ensure top placement on the DPS meter is regarded as dangerous and could lead to a raid wipe:

> *“Sometimes they’ll go away with blaring DPS, doing as much as they can to show off, and come top of the DPS meter on everyone’s screen. But they don’t care about the tactics for the boss, or care what their specific role should be, they just want to show off and be top of the DPS meter. Which should be a secondary goal, not a main goal.”* [P2, 23, F]

7.5.3.2 Meter Content as Limited in Scope

While the information represented in meters is useful during raids and is used to multiple ends, issues can arise with the way in which this information is interpreted. In particular, meter content can be misleading because it is limited in scope and does not capture all of the actions that could be filed within an individual’s ‘contribution’. We found several distinctions on this matter, most of which can be linked to the aggregation of actions.

First, interviewees recounted experiences related to the way in which their contributions had been *misinterpreted* by other group members. This is related to the fact that meters are poor at capturing peripheral and supporting contributions that are not related to damage or healing but are nonetheless critical to the group. For example, we saw earlier that certain classes have leeway to disappear from the meters when their role demands it, as when performing ‘crowd control’ on additional enemies. The fact that meters do not immediately provide evidence of these contributions can make it seem like certain individuals are doing little work when, in reality, their contribution is critical for success:

> *“My role was, as a hunter, I was used for a lot of the add work. Misdirecting. Which means that you’re switching off the main target to control something else and misdirect it... it brings your DPS right down, so there’s no way you’re going to be top of the meters.”* [P4, 46, F]

An additional issue related to misinterpretation is that certain ‘hybrid’ classes blur the boundaries between ludic roles, i.e. their contributions are spread across both damage and healing meters. This means that hybrids are often placed relatively low on both meters, which, at first blush, could imply that they are contributing very little or are perhaps not as good in their defined role as those above them.

Second, simple aggregations of actions can potentially *misrepresent* the value of individual contributions because information about the *quality* and *relevance* of actions is unavailable. For example, a player may be using certain abilities that are captured by the meter but are ineffective for the task at hand—an example might be the use of abilities to which the group’s enemies are immune. Meters do not allow players to detect such behaviour because they simply aggregate performance without accounting for contextual details. Furthermore, meters can cause certain contributions to be *misattributed*. This occurs when one group member's actions have knock-on effects that benefit the group as a whole but suppress the importance of the original contribution. For instance, one character class can bestow others with a 'blessing' that gives an extra 10% to each character's strength. However, the provider of the blessing receives no credit for this contribution in the meters, whereas the outputs of everyone else are inflated by the strength increase. While it is important to recognise that groups in WoW are typically aware of these peripheral contributions, taking meters out of context could make it appear as though one person has contributed more to the group when, in reality, their contribution was absolutely dependent on the involvement of others.

Lastly, the value of damage meters as measures of contribution can vary according to features of specific encounters. This relates to the fact that different classes within WoW can be more or less effective during certain fights. For example, there are certain enemies that are more susceptible to ranged magical damage, and thus Mages, Warlocks and Priests are at an advantage during these fights. In other cases, however, these classes cannot fully contribute because many of their most powerful abilities require the character to remain stationary—this is problematic during encounters where raiders must make continual movements to avoid threats in the nearby environment:

> *"If you had to move a lot, casters would generally be lower down because they're moving so much whereas if you're tanking or if you're not ranged then it's easier, so there're differences depending on what the fight is."* [P5, 18, M]

The point here is that, as simple aggregations, the contents of meters do not account for subtleties of different encounters. This means that some raid members are at an inherent disadvantage, but meters fail to portray this, making it seem as though some players are underperforming relative to others when they in fact cannot maximise their own performance. Of course, competent raid groups will be aware of these issues, but one interviewee recounted an example where a group leader had berated her for seemingly poor performance. In this case, there was a lack of common ground regarding her character's abilities and the structure of the encounter:

> *"The first boss in Icecrown, he does an attack that puts you in a spike if you're too far away... so everybody has to run into the boss, apart from hunters because you've got an area that's a deadzone where we can't attack, so we have to stand outside of the deadzone, but that means we get spiked... And we had one of our officers get rather antsy with me a couple of times for getting spiked and not DPSing... it took our class leader to pull him and say, look, learn the class, you have no idea what you're talking about."* [P4, 46, F]

Taken together, these issues mean that it can sometimes be difficult to get a good sense of performance based on meters alone, primarily due to their limited scope as measures of *overall* contribution. This is especially true when efforts are undermined by aspects of an encounter that are beyond a player's control.

7.5.3.3 Amplifying Evaluation Apprehension

Our final theme pertains to evaluation apprehension—the fear that contributions will be viewed negatively by other members of a group, in turn causing individuals to withhold such contributions to avoid criticism (Diehl & Stroebe, 1987). In the context of WoW, the fact that contributions are aggregated and regarded as a measure of 'performance' may cause individuals to be nervous about the contributions they make and the way in which those efforts are interpreted by others. Such an experience was recounted by one interviewee who, as a new raid member, had been worried about performing poorly and being evaluated in a negative light. (Raids often require new members to undergo a trial period where performance is analysed and a decision is made on whether the individual can stay.) Her response hinted at an almost self-fulfilling prophecy, where, because she was nervous about being evaluated, her performance deteriorated below her usual standard:

> *"It made me feel anxious before, like I'm being scrutinized. I was worried about how people would judge my actions, and that made me worse in some ways."* [P2, 23, F]

Evaluation apprehension is likely true of group participation irrespective of the work context, yet meters may amplify the effect by publicising detailed aspects about individual performance. Any negative behavioural outcomes may therefore be especially problematic in group settings where each individual must meet an acceptable level of performance, as was the case in WoW raid groups.

7.6 Discussion

The present study aimed to consider how meters in WoW raiding allow players to make inferences about contributions and, by extension, the efforts of other raid members. Our study sheds light on a range of beneficial aspects associated with meters, but also unveiled a series of sociotechnical concerns and corollaries that suggests meters can, at times, be problematic. These issues may have implications for the use of contribution meters to assess and manage fairness in other contexts. In the present discussion, we will first consider our findings in terms of how meters support awareness in WoW raid groups. We then consider ways in which game designers might productively elaborate on existing meters to alleviate certain problems made salient by our analysis. Finally, we will consider how contribution meters, as a general design idea, could be applied to other work-related scenarios in CSCW.

7.6.1 Meters and Awareness in World of Warcraft

At a broad level, our results emphasize the critical importance of awareness during WoW raids, and fall into line with contentions about the importance of awareness from elsewhere in CSCW (e.g. Schmidt, 2002; Gutwin & Greenberg, 2002). If players do not maintain ongoing awareness of what is going on in the raid, the risks of individual death and collective failure appear to be heightened. Awareness appears to operate at several levels. First, a general sense of 'raid awareness' seems important—players must continually update their knowledge about the interaction between themselves, emergent threats, and the raid's primary target. Contribution meters assist players in developing this awareness by offering clues about who is active and what their overall status is. However, it is important to recognise that meters are *complementary* to existing communicational and coordinative practices—most raiders also use voice chat programs to issue commands and share information. Thus, if a player does disappear from the meters, players can use other channels to request verbal updates from the person or persons in question about their non-participation. Additionally, meters support various lower-level aspects of raiding. These aspects range from allowing players to assess individual progress over time to ensuring that contributions meet the standards required for a given encounter. These findings, combined with others in our results, help to throw light on the uses of awareness information within WoW, in turn providing an improved understanding of individual and team behaviour in collaborating raid groups.

Our initial motivation for studying awareness in WoW was to assess the extent to which meters are useful for supporting judgements about fairness. Several findings speak to the utility of meters in this regard. First, raids are better off with a certain degree of equality in contribution—groups that are capable of achieving roughly equal outputs from their members minimise the risk of failure that can result from the death of individual characters. Meters help raids strive towards this equality by displaying a clear threshold for performance and, in line with our contentions, by allowing players to compare the performance among team members, both publicly and privately. Additionally, meters allow raid leaders to identify underperformance, which is in turn dealt with as appropriate. As hinted by our responses, meters also help raid leaders to ensure individuals are contributing in line with expectations about fairness; thus, just as in other group settings, there is some basic expectation that everyone must contribute 'something' and that free-riding behaviour is unacceptable.

However, we found a range of sociotechnical issues that imply differential benefits of meters as measures of contribution. To recap, our results suggest that meters can:

- emphasize competition over coordination, especially if there is status associated with being high on the meters.
- obscure peripheral and supporting contributions that underpin task progress.
- fail to inform the reader about whether or not particular contributions are appropriate.
- distort the value of some contributions while suppressing the importance of others.
- amplify the feeling of evaluation apprehension for those new to a group.

While these side-effects are necessarily specific to WoW, in terms of the sorts of behaviours they promote or discourage, it is possible that similar issues could arise if contribution meters were applied elsewhere. First we will consider the problems specific to WoW so as to offer resolutions relevant to game designers. Then we will consider the significance of these issues in light of our design goal of supporting fairness.

7.6.2 Implications for Awareness Support in Collaborative Games

Our results initially imply some differential benefits of meters; while they are useful for awareness, they can cause unintended side-effects which may affect players' enjoyment of raiding in WoW. This is potentially undesirable from a user experience perspective. However, it is important to recognise that some of these issues are probably negated through groups' appreciation of context—even though individuals with supporting roles may be relatively low on the damage meters, it is not the case that these players are ejected from groups. Rather, the issue seems to be one of ensuring that players are appropriately accredited for individual actions. This seems especially important if collective outputs factor into reward decisions, as sometimes occurs during the assignment of loot in raiding groups.

One problem is that aggregation of a single job function, as embodied by damage or healing meters, fails to account for supporting contributions. We suggest that the general idea of 'meters' is something upon which game designers could elaborate. One possibility might be a more general contribution meter that assigns a uniform rating based on the use of different actions over time, rather than a flat measure of a single output. For example, the use of crowd control abilities or other supporting behaviours could be factored into an overall rating so that each player feels that their contribution has been acknowledged. Alternatively, a meter could be built that suggests an estimated parameter for individual performance based on the characteristics of a given encounter (cf. Torkel, 2007). A player's contribution could then be gauged relative to this parameter, with points awarded or deducted for falling above or below the performance threshold. The challenge for designers will lie in determining how different actions should contribute to such a rating. These considerations, as well as the issues identified within this study, are likely be of clear relevance for interface designers who wish to support both awareness and harmonious working in collaborative gaming experiences.

7.6.3 Implications for Contribution Awareness

Our findings suggest that meters, as one potential way of implementing contribution awareness, do allow for fairness judgements and help raid groups to negotiate assignment of rewards. Meters also provide a range of awareness-related benefits, e.g. in monitoring personal development, and it is possible that similarly beneficial aspects might emerge if meters were applied in other contexts. However, we must remember that meters are one way, and not the *only* way, in which contribution awareness could be implemented—we regard them as representative of one point in the design space for tracking and visualising contributions. Below we outline some more general lessons and concerns related to meters and the broader idea of contribution awareness.

In detailing our design idea, we touched on the fundamental tradeoff between awareness and privacy. Privacy is not something we asked about specifically in this study, but is nevertheless important because the extent to which a system infringes privacy can impact its overall acceptance (Plowman *et al*., 1995). Thus, any other system that offered detailed awareness information might make its users feel uncomfortable. One interesting aspect of the present study is that, even though we asked players about the negative consequences of meters, the issue of privacy did not arise in our results. Players do not seem to mind the presence of meters, which might be regarded as surprising in light of findings from the literature on 'electronic performance monitoring and control systems'—tools that allow companies to collect, store, and analyze the performance of employees (Alge, 2001). Such systems are similar to meters but have been known to cause discomfort and stress among workers due to perceptions about monitoring (Alge, 2001). One possibility is that the groups we studied are unconcerned about privacy because meters are an accepted, and expected, aspect of the raiding experience. Additionally, meters are not merely used by the 'higher-ups' to monitor subordinates; rather, displaying information about individual efforts has a clear benefit to the rest of the group. The general lesson here might be that, if a contribution meter were implemented in another setting, the potential privacy issues associated with increased information disclosure might be negated by ensuring that there is a clear collective benefit and that monitoring is not used for its own sake. The acceptance of detailed awareness tools might also depend on the extent to which equality in contributions is necessary for success, as was the case in the groups we studied.

The present study highlighted the potential for unintended side-effects resulting from public visibility of contributions. It is obviously quite difficult to predict whether any similar effects would emerge if contribution meters were applied to other tasks and systems, but there are perhaps some broader lessons that can we consider. The first concerns the potential influence on group goals. We saw that contribution meters could shift the focus of some individuals away from the primary task and towards secondary goals, e.g. competition, that were less appropriate for coordinated task completion. Such an effect is undesirable and represents something of an unwanted risk from contribution meters. Perhaps designers could find some way of keeping a group's focus on the primary task, e.g. through behavioural nudges or evocation of particular normative standards (cf. Thaler & Sunstein, 2008). Another issue concerned the limited scope of what was displayed in the meters we studied. Damage and healing, as two measures of activity, were critical for groups' success, but they were not the only actions important for task progress. Displaying only one aspect of an individual's job role meant that certain players appeared to contribute less than others. The basic implication here is that designers will need to be careful in choosing which contributions to include, as well as the way in which those contributions are represented. Simple aggregations of action may not be enough—our earlier contentions about damage meters versus more general measures of 'contribution' speak to this need.

7.6.3.1 Broader Design Challenges

In terms of implementing contribution awareness in other systems, there are at least two broader design problems worthy of consideration. One pertains to the issue of how different contribution types might be usefully aggregated into a readable metric, and indeed whether different contributions can be usefully compared at all. Assessing fairness through comparison of contributions could be particularly difficult when subcomponents of larger projects are inequivalent. In Chapter 3 of this thesis, we studied groups working on their Integrated Project coursework, a task that requires writing, computer programming, and software design, as well as intangible contributions associated with project management. Comparison of these subcomponents may lack meaning because it is hard to gauge what each is worth relative to the next. Perhaps measures like percentage contributions or time on task, as opposed to aggregation of actions, would be more appropriate—this is something that can only be explored with further work.

A second design problem pertains to the issue of work *quality*. Our arguments about contribution awareness are based on the premise of making people aware of actions, and meters enable this idea by collapsing contributions into one readable metric. However, the issue of contribution quality is something that meters were seen to obscure. While it is true that 'quality' can sometimes roughly equate to overall output—as with the amount of damage produced by a players in WoW—it can also be measured in terms of appropriate actions. This highlights a deficiency in the basic blueprint of meters, in that aggregation of every action irrespective of usefulness could be misleading when applied to other contexts. This might then be problematic for judging fairness, especially if quality were a critical criterion for success. Of course, contribution meters could still support some degree of fairness by providing a persistent account of who contributed what. In tasks where quality is more coarsely defined, identifying 'good' from 'bad' is something that may be technically unfeasible and simply has to be left up to the group. These are very much the sorts of considerations we aimed to highlight through the present study.

Perhaps the next step for exploring contribution awareness would be to investigate how meters, or other representations, can be used to publicise contributions in other tasks. We do not propose specific designs here, but there are many cases in which improved contribution awareness could prove useful. Collaborative document editing is one such example; although many modern applications have commit logs, edit histories, and tracking of changes (each of which could conceivably be used to gauge contributions) meters could provide a more concerted visualisation of each person's contributions to the work. Such a visualisation might prove useful for appropriately accrediting authors for their work. The literature on academic authorship suggests that 'ghost authoring' and coerced accreditation are pervasive (e.g. Flanagin *et al*., 1998) and meters could be used as evidence to help determine author order. However, this idea touches on our earlier concerns abour *how* contributions can be aggregated, and indeed whether or not detailed information would even be desirable. Writing involves not only adding text, but also requires rewriting, organising content, correcting typographical and grammatical errors, etc. These aspects could be difficult to represent in a single metric, perhaps warranting the need for several that gauge key outputs. The question of whether collaborators would even *want* such information is an open one. Such considerations only imply the need for further work—we suggest further explorations of the design space and evaluations of other mechanisms that aim to support fairness through contribution awareness.

7.6.4 Study Limitations & Future Work

This work is not without its limitations. Perhaps the foremost limitation is that our interviewees can not be considered as representative of the general WoW population. Although we acknowledged this limitation earlier, it means that we can not make definitive statements about the importance or prevalence of one issue over another. While our dataset clearly captured the more salient issues, there may also be other practices related to meters that were not surveyed in this study.

In limiting our focus to raid groups, this study neglects a specific subset of grouping in WoW: 5-man instance runs, which we suspect are actually more prevalent than raid groups. It is much less effortful to form a 5-man group, and the dungeons such groups encounter usually have a much lower threshold for coordination and time commitment than do raids. While our focus on raids was motivated by their intrinsically collaborative nature and by our interest in contribution meters, future work could explore smaller groups to see if any of our contentions could be refuted or adjusted based on the work of smaller groups.

Another aspect of WoW not covered in this study is player-versus-player (PVP) grouping. In these settings, players compete in teams of between 2-5 to defeat other players. The emphasis is placed on fighting other humans rather than in defeating computer-generated challenges. It is unclear whether contribution meters are relevant to these tasks. Our findings might differ to what is required by players in such competitive circumstances.

While we have aimed to consider how metrics support grouping, one iteration on this study would be to employ quantitative approaches to further probe fairness in raid groups. For instance, it would be interesting to explore whether raid groups featuring more equitable contributions are more satisfied or successful than those where contributions are more polarised. Such a study could use server logs to track raiding progress over time (cf. Bardzell *et al.*, 2012) and then compare progress with the contents of damage and healing meters at particular intervals. This would provide further insight into the relative importance of equity in raiding contexts.

Lastly, the results of the present study, as well as our contentions about contribution awareness, pertained to synchronous work. We believe that the idea of contribution meters could easily be applied to work done asynchronously, but the updated aggregation of work products could prove problematic for interpreting contributions over time. We believe this is an issue that warrants further work.

7.7 Chapter Summary and Conclusion

This chapter has aimed to consider how collaborators could be supported in making accurate judgements about fairness in contributions. We first considered the importance of awareness in CSCW and the inherent difficulty of judging fairness within collaborative systems where details about actions are sometimes obfuscated. We suggested that awareness mechanisms could support fairness by providing detail about individual contributions and by allowing collaborators to engage in social comparisons so as to observe deviations from equality. In line with these contentions, we explored the use of existing metrics for judging contributions within World of Warcraft raid groups. Our analysis revealed a range of potential benefits, but also highlighted some sociotechnical issues that can result from use of meters. We suggest that the design ideas embodied by meters could be applied in other collaborative tools that aim to support fairness preferences.

The present chapter has several contributions, the first of which is our design suggestion of 'contribution awareness' as a means of supporting fairness in collaborative systems. Second, we have elicited an improved understanding of the utility and importance of awareness in WoW raid encounters. This understanding might also be extensible to other collaborative grouping situations, and provides a theoretical contribution to the literature on virtual worlds. Third, our analysis of meters implies an opportunity for game designers to expand existing metrics to provide broader measures of 'contribution' that are not confined to a single job role but are instead reflective of a broader range of contributions. Such a meter would offer improved accountability for peripheral and supporting contributions. Finally, we considered the extent to which contribution meters could be applied in other settings—of course, further work is necessary in this regard. Nevertheless, we believe that visualisations of contributions relative to one another may be a useful step on the the path to improving individual accountability and allowing groups to judge whether work is being completed fairly. The following chapter considers these contributions alongside those of earlier chapters in attempt to offer conclusions and a roadmap for future work.

Chapter 8

Conclusion

8.1 Thesis Review

The central aim of this thesis has been to investigate the subjects of fairness and division of labour during collaborative work, focusing in particular on computer-supported collaboration. We aimed to explore the issue of how fairness might be relevant to collaboration, as well as the extent to which people strive for fairness during work assignments. Our studies on this matter highlighted the guiding role of fairness in division of labour, as well as the apparent difficulties of gauging fairness during computer-mediated work. Based on this latter premise, we explored one way in which computer systems might allow collaborators to make more accurate judgements about fairness during the completion of their work.

This thesis was scoped to consider fairness in the division and completion of collaborative work. In Chapter 2, we aimed to establish conceptual foundations while drawing a link between collaboration, fairness, and CSCW. We sought to consider appropriate definitions of both collaborative work and fairness, and considered concepts related to each. We then attempted to explore the relevance of fairness to CSCW through review of prior work. We found that fairness has not yet received an in-depth consideration within CSCW and HCI more generally, save a few specific cases where the concept has been mentioned as a matter of relatively low importance.

In Chapter 3, we conducted a survey of small workgroups to explore the link between fairness in division and completion of labour and overall satisfaction with a collaborative project. We found that perceived fairness in the division, and completion, of work was correlated with three measures of satisfaction (process, products, and performance of colleagues). Correlations were positive and of moderate strength, and although our approach was mostly exploratory, the findings are suggestive of a relationship between perceived fairness and overall satisfaction. Participants' statements were also revealing about the struggles of non-participation and free-riding, and implied that it may be difficult to assess fairness due to the absence of information about individual and collective contributions.

In Chapter 4 we introduced a more immediate approach to studying fairness in the shape of an empirical model based on the classic ultimatum game, an experimental framing of negotiation that has been used to study fairness in hundreds of prior studies. We proposed a novel transfiguration of the ultimatum game such that, rather than bargain over a financial payoff, participants instead bargain over a potential reduction in workload. One player makes a proposal about how to divide the work, and the other chooses to accept or reject. In the event of acceptance, players complete the game as agreed. In the event of rejection, players work alone and do not collaborate. We offered our model, the division of labour ultimatum game (DLUG), as an explicit framing of the planning phase of collaborative work. This model is then linked to the execution of work by the fact that participants have to complete the task in line with the outcome of the game. We proposed to explore our model within this thesis and, to this end, conducted a thorough review of empirical literature relevant to the ultimatum game. Our review helped to ensure that our own experiments were built on the strongest possible foundations. We also considered prior use of economic games within the field of HCI. This allowed us to identify ways in which our model could be used to elaborate on earlier suppositions, in turn offering meaningful contributions to HCI, CSCW, and the wider social sciences more generally.

Chapter 5 explored our model in a series of empirical studies. We opted to scope our investigations by focusing on the subject of synchronous collaborative information seeking, an emerging area of research in which division of labour is a relevant research problem. Four studies were reported: thesis study 2 offered an exploration of the DLUG in its most basic form, and study 3 explored the extent to which a more or less attractive search task might impact allocations of labour. Study 4 explored the threat of rejection as a potential explanation for behaviour in the DLUG, and study 5 explored allocations in the absence of an explicit allocation procedure. We found a consistent preference for fairness, characterised by equality in distributive allocations of workload. Based on our analysis of task completion times, we also found evidence to suggest that participants strived towards achieving equality in the completion of their work. Our discussion of these results was guided by relevant literature; we argued that fairness in initial allocation of workloads might be the result of rapid decision making that conforms to social norms while providing a strong cooperative signal, perhaps implying collaboration readiness. We considered whether fairness in completion might be tapping so-called 'effort-matching' effects identified elsewhere in social psychology. Finally, we identified strategies used to coordinate division of labour and prevent redundancy during the search task. Such strategies have immediate implications for the design of collaborative search systems.

In Chapter 6, study 6 explored the extent to which division of labour and fairness come into play during the execution of real-world collaborative search tasks. Additionally, we used two existing systems to explore real-world collaborative search behaviour and examined the extent to which our chosen systems were appropriate for supporting everyday tasks. Interviews held with participants revealed that, at a very general level, the two systems were regarded as useful, but collaborators identified a range of problems related to sensemaking, information overload, and awareness. These problems present immediate implications for design. Regarding fairness, we found that people were less worried about policing the equality of contributions in these casual, everyday search tasks, and that there was a variable perception about the importance of fairness and the work itself. In some pairs, apparent unfairness was tolerated to allow for task completion. However, participants did expect their collaborators to at least contribute to some meaningful standard, and monitoring said contributions was sometimes regarded as problematic due to limited awareness. This latter finding can be likened to suppositions made elsewhere in the literature about the difficulty of monitoring collective contributions (Galegher & Kraut, 1990; Gutwin & Greenberg, 2002; Hertel *et al.*, 2003), and runs parallel with findings from study 1, where students regarded it as difficult to judge fairness due to the ephemeral nature of offline work.

In Chapter 7, we drew on our earlier findings to explore how groups might be allowed to make judgements about fairness during computer-mediated work. We proposed the idea of *contribution awareness*, arguing that, to enable more accurate judgements about fairness, systems could provide detailed information about the ongoing contributions of collaborators. Furthermore, we suggested that systems should allow for social comparison to facilitate judgements about relative equity in contributions. Study 7 explored the potential of our idea, where we examined the use of contribution meters by groups in the online game World of Warcraft. We studied the extent to which meters, as an example of how contribution awareness could be implemented, support judgements about fairness. We also offered insight as to how awareness supports raid groups more generally. Our study suggested that awareness of contributions can be appropriate for judging fairness, but raised four potentially negative consequences that suggest designing for fairness might not be straightforward. We considered the extent to which these issues might apply to other settings.

In this final chapter we draw the thesis to a close by considering broader implications. We also identify limitations and opportunities for future work based on the present research.

8.2 Integration and Implications of Thesis Findings

In this section we aim to integrate findings from across the thesis to provide new perspectives on fairness in CSCW. Our interpretations are positioned relative to the research questions posed in Chapter 1. This helps to structure our arguments while bringing the thesis work full circle.

We began this thesis by proposing three questions that were intended to frame our studies on fairness in collaborative tasks. We can now draw on our study results to consider initial answers to these questions and tease out some of the central themes that have been made salient through our thesis research. In particular, we identify three themes:

- First, fairness can be a critical requirement for harmonious working in some collaborative situations.
- Second, context effects mean that unfairness is tolerated in other circumstances.
- Third, designing for fairness is a legitimate goal for CSCW, but design for fairness may be tricky to the extent that there may be occasions where it is better *not* to design for fairness.

Each of these statements is elaborated and positioned relative to our research questions in the following sections. In addition we offer a model of division of labour that specifies the questions an individual might ask when assessing and monitoring fairness during collaboration. The model may help theorists to reason about how fairness is realised at different stages of the collaborative process, and might assist practitioners in thinking about interventions to support fairness.

8.2.1 Theme 1: Fairness can be Critical in Collaborative Work

Research Question 1 was: **What is the relevance of fairness for collaboration and collaborative systems?** Having scoped the thesis to consider division of labour, our findings imply that fairness is strongly relevant to both the division and completion of collaborative work. In thesis study 1, for example, we found that fairness can be at the forefront of individual concerns throughout the collaborative work process. Members of student teams appeared to be more satisfied with the work process when they perceived that fairness in division of labour was high, and unfairness was a source of dissatisfaction for those who incurred free-riding. The implication here is that failure to comply with fairness norms may lead to poorer social health for a group. It is not clear whether this deficit would cause groups to underperform but, based on the related literature, it seems sensible to suggest that performance would be hampered when certain individuals free-ride (e.g. Brooks & Ammons, 2003). Since fairness is a salient emotional concern during collaboration, supporting fairness might be seen as a design goal for any CSCW system that mediates the assignment and completion of work-related tasks.

Turning to Research Question 2, we asked: **To what extent do people strive for fairness in the division of labour?** Findings from this thesis indicate that there is a preference for allocating work in line with fairness norms. Fairness also affects behaviour during the enactment of work, whether that means adhering to initial commitments or in adjusting workloads to meet emergent challenges. The DLUG experiments reported in Chapter 5 provide strong evidence of both of these phenomena. In addition, ensuring fairness in the work process may partly be about pacing workrate so as to acquire overall equity. The yoking of task completion times can be viewed in terms of compliance with fairness norms (cf. Jackson & Harkins, 1985). Collaborators may adjust their efforts to ensure equality in the overall outcome of the work process and this normative standard of workrate may then be retained for future application in work settings (Kelly & McGrath, 1985). However, this research exposes important theoretical considerations for fairness in division of labour. One is that a fair allocation does not always involve a direct mapping to equality. That is, it may not be fair to divide a task evenly if other factors would provide rationale for legitimate deviation from equality. Groups may invoke other allocation rules in such circumstances. An example of such a rule would be ‘divide by competency’, as evidenced by the assignment protocols used by some groups in study 1. In such cases, individuals considered it fairer to match tasks to relevant skillsets and individual strengths, rather than attempt to ensure identical quantitative workloads. Not only is this putatively fairer by not requiring people to handle tasks with which they have no competency, it also saves time in negating the need to learn new skillsets. Groups must therefore balance the subject of fairness with other issues such as *time* and *efficiency*, but further work is required to understand how these considerations play out in the real world. Other examples of efficiency-related behaviour were evident in some of the DLUG experiments, where some individuals kept more work for themselves (instead of delegating) as they believed that retaining control would allow for quicker task completion. Thus, strict quantitative equality can be suppressed by allocation rules that bring efficiency to the fore while still accounting for fairness preferences.

8.2.2 Theme 2: Context Affects Tolerance of Unfairness

Implicit in the preceding discussion is the idea that fairness is critical for harmonious working and that this requirement may be uniform across a wide array of work settings. Such an argument is called into question by study 6 of this thesis. In that study, collaborations occured between friends and family, and although most wanted to work in line with equality, unfairness was often tolerated in favour of task completion. Thus, a second theoretical implication is that the importance of fairness is not universal; context effects must play a role in determining the extent to which people care about and 'police' fairness preferences.

Based on our results, as well as arguments in the literature (e.g Leventhal, 1980), we suggest that the extent to which people enforce their preference for fairness is variable depending on the work situation, the type of work that is being done, and the people with whom one is working. The combined influence of these factors is likely to affect the salience of particular rules and norms that guide the collaboration and completion of work. Moreover, effects associated with context may provide rationale that is sufficiently strong to allow for tolerance of unfairness. To consider this further, it is worth dwelling on how the contexts studied in this thesis differ.

Studies 1 and 7 looked at existing work situations in which team members had interdependence and were asked to make contributions to achieve some overall shared goal. Studies 2–5 likely invoked work-related norms that would be similar to the settings of studies 1 and 7. (We see our DLUG as a tool that exposes the way in which people think about real-world division of labour and collaboration, just as the UG can be considered as invoking norms centred around bargaining and exchange). We characterise these and other task-focused settings as *formal* contexts in which expectations about participation are quite specific. Conversely, we characterise the settings of thesis study 6 as an *informal* work context. We hypothesize that this informality affects the extent to which compliance with fairness norms is critical for harmonious working. As a brief elaboration, we can tease apart the extent to which these different settings might emphasise different rules for evaluating fairness. We delineate the two settings in terms of the following properties:

- *Rules*: Workplaces are settings that exist within organisational frameworks that dictate what is and is not permissable (Holt & Thaulow, 1996). Although tacit rules govern all aspects of human behaviour, we argue that informal settings are not governed by clearly articulated organisational rules but are instead guided only by those that have been negotiated or established over time within the confines of the relationship.

- *Rewards*: Formal contexts, such as professional workplaces, involve work that is done in pursuit of material reward. Informal work can be regarded as a discretionary activity for which there is no material reward.

- *Roles*: Formal settings involve roles and responsibilities, as evidenced by job titles and management hierarchies, that are less evident in informal settings. While personal relationships do promote roles to organise certain tasks, e.g. division of labour in the home (Mikula, 1998), the structures of power are less realised and thus people are not typically held accountable to the kinds of expectations about their role as they might be in formal workplaces.

- *Expectations*: Workplace relationships exist primarily to serve the needs of the employer and to complete professional endeavours. Such relationships will be governed by expectations about work and justice rules, such as the expectation that one will do one's duty and comply with relevant norms, i.e. fairness. The relationships that exist in informal settings—such as those between family, friends or romantic partners—are not formed on the basis of task completion but instead exist as a means of satisfying mutual social, moral or hedonic needs. There is likely to be a set of expectations about fairness but they are likely to be oriented towards goals that are not task-oriented.

- *Retribution*: Formal and informal settings differ in terms of the extent to which sanctions are imposed if one fails to behave in line with expected rules. In a workplace, failure to perform to an expected standard (i.e. to contribute fairly) could result in penalties ranging from docking of credit (as would occur in student workgroups) to the loss of one's job. Non-participation in the type of work that was being done in the non-professional settings is unlikely to lead to such sanctions.

These considerations are certainly not exhaustive but at least serve to articulate some of the factors that would affect the extent to which people enforce fairness norms. Moreover, they help to characterise the sorts of situations in which unfairness might be socially acceptable, further emphasising the fact that fairness is not unidimensional—not only do granular perceptions about fairness differ from person to person, the weight that is assigned to particular rules will differ between contexts. People are known to selectively apply and adjust fairness rules depending on the circumstances at hand. For example, Leventhal (1980) describes how justice rules are weighted in accordance with the social situation. More weight is likely to be applied to the evaluation and comparison of contributions "in settings in which productivity and task achievement are the primary concern" (p. 10), whereas people assign weight to other rules when group solidarity and maintaining harmonious relationships are of greater importance. We suggest that the informal contexts captured in study 6 fit this latter category—although there was undoubtedly some emphasis on getting the job done, task completion was not to be emphasized over maintaining a harmonious relationship. And although it could be argued that fairness itself would help to maintain a good relationship, perhaps there were issues beyond the scope of the study that warranted legitimate deviations from equality in the collaborative search task. (Such as one individual doing more work in other areas of the relationship).

In addition, studies from the UG literature also emphasize that *intentions* guide conformity to equality. For example, Nelson (2002) used a truncated ultimatum where the maximum possible offer was capped at 20% of the pie (recall that offers of 20% or less are typically rejected in ultimatum games). The main finding was that restricting the maximum possible offer to $4 (20%) significantly increased the number of participants who were willing to accept the offer, from 17 out of 44 participants in the standard game to 42 out of 44 in the truncated version. In this case, responders know that the proposer has made a best effort and tried to be as fair as possible in light of the circumstances; thus, it is often 'the thought that counts' when evaluating fairness. In the informal settings of study 6, it may be that participants knew their partners were not contributing equally but that this lack of contribution was not a result of laziness, free-riding, or spite. Individuals might not have been unfair deliberately but were instead distracted by other concerns, and this knowledge was enough to relax policing of fairness norms.

In sum, it is clear that fairness rules may be intruded by other variables, e.g. the type of relationship that exists between the persons involved. Fairness rules are not hard and fast, and appear to be applied selectively depending on salient characteristics of the situation. This may make fairness a tricky design goal—not only is fairness malleable but it is often difficult to predict the exact circumstances in which a given digital technology will be used. This means that any design decision oriented towards supporting fairness might be evaluated differently from one situation to the next. What we must now consider are the implications of the preceding discussion, as well as our other results, in terms of how they might impact the design of CSCW systems.

8.2.3 Theme 3: Fairness as a Tricky Design Goal for Collaborative Systems

Our final theme can be positioned relative to Research Question 3, which asked: **How might collaborative systems be designed to account for fairness preferences in division of labour?** In answering this question, Chapter 7 saw us put forth two propositions. First, we suggested that designers could consider providing interface elements that display sufficiently detailed information about the ongoing contributions of each person within a collaborative team. This suggestion arises from the fact that, when making assessments about fairness, an immediate challenge concerns the collection of information necessary to make a judgement in the first place. In offline groups, fairness is naturally quite hard to assess because actions are fleeting and ephemeral. In distributed or computer-mediated work, judgements might be even more taxing because work is rendered invisible and awareness of colleagues can be low. This means that fairness judgements, irrespective of the work situation, are always somewhat hazy—people can only assess fairness based on what they *believe* about a situation. Digital tools have the power to strengthen collaborators' judgements of fairness by providing people with records of action that would not normally be available. Through proper design, information about collective action could therefore allow for fairness judgements that are more closely aligned with the realities of a work situation.

Our second proposition emphasised that fairness is assessed through *social comparison* (e.g. Cohn *et al*., 2009). Evidence from thesis study 7 demonstrates the potential benefits of allowing comparison of contributions. In our example context (World of Warcraft) players use meters that visualise each person's contributions to ongoing collaborative tasks. Meters allow players to make assessments about their contributions relative to those of the group, allowing for monitoring of individual and collective performance over time. Meters are also used to mediate the distribution of group rewards and ensure that individuals are appropriately compensated for their efforts. However, we also saw that meters could prove potentially problematic because their scope is relatively limited, meaning that certain contributions were obscured or missing altogether. These consequences raise important questions for thinking about how designers can tailor CSCW platforms to support fairness. We now detail some especially salient questions and offer a short discussion of each, before proposing some preliminary guidelines that designers can follow when thinking about fairness in collaborative systems.

1. *What should be displayed in a tool designed to support fairness?* The answer to this question will, of course, depend partly on the nature of the task at hand. In collaborative writing, for example, it would be prudent to select metrics that are relevant to the task of creating and editing text. But the broader challenge is unveiled by deeper consideration of the language we used in study 7, where we spoke of meters showing 'actions', 'efforts', and 'contributions' as if they were interchangeable. Further reflection suggests that each could be considered as conceptually distinct from the next. *Actions* might refer to specific detail about the operations an individual is performing or has previously completed. *Effort* might describe the intensity with which those actions were performed, or might be assessed via a measure derived from actions. (For example, 'actions per minute' is a common measure of workrate in some online games). Finally, *Contributions* might be positioned as an aggregation of an individual's actions and efforts. In WoW meters (see Chapter 7, Fig. 7.2) contributions are displayed in absolute and percentage terms. A designer might choose to display both of these, or indeed neither if some other metric is more suitable. Fairness could be assessed in terms of completed subtasks, percentage of the group's overall efforts, or by an entirely different method. These considerations highlight the fact that selecting and displaying the right information is an ongoing challenge, and its solution will likely differ across a swathe of different collaborative situations.

2. *How is information intended to support fairness used across different settings?* We already know from the literature and the results of this thesis that fairness is not unidimensional. This means that it will probably be very tricky to reach consistent guidelines about how and when designers should aim to support fairness. If we think about a concept like *awareness*, its general nature remains fairly consistent across settings—fundamentally, it is about knowledge of what is happening in a workspace, who is around, where objects are, who is using them, and so on (Gutwin & Greenberg, 2002). This means that something intended to support awareness is likely to be successful when applied across different collaborative contexts. As an example, interface widgets that display the current location of colleagues are useful in many different CSCW platforms, ranging from chat rooms like Babble (Erickson *et al.*, 2002) to online games like Warcraft. Moreover, support

for awareness is relevant to any collaborative context where information about colleagues would improve one's understanding of shared action and the work process. Since we know that fairness is of variable importance across different settings, it is possible that information about contributions would be extremely useful in one setting but infuriating in another. However, our understanding of how those contexts differ, beyond the extent to which they are governed by formal rules, is limited at this time. We see fairness as quite different to the phenomena with which CSCW is used to dealing. Compared to other concepts that underpin collaborative work—conversation, cooperation, coordination—fairness is more malleable and and subject to change; it has temporal qualities and differs in importance from one situation to the next. Only future research can help to remove some of these ambiguities from our current understanding.

3. *Should we even design for fairness at all?* We saw in our study of WoW that the information used to gauge fairness had a number of detrimental consequences for group interaction. Some interviewees described how metered content about contributions could cause players to focus on competition over cooperation. Others described how the limitations of meters meant that certain contributions were obscured or missing altogether. Further still, the value of contributions could be misinterpreted, or misattributed to the wrong individuals. Not only do these findings specify the limitations of meters as tools for evaluating fairness, they are enough to suggest that designing for fairness is tricky and might lead to unanticipated consequences. Our ideas about supporting fairness are underpinned by the concept of monitoring collective actions, but there may be times when monitoring is distracting or is costly to the extent that it causes the pace of groupwork to slow down. Perhaps encouraging people *not* to reflect, and not to ask questions like "how am I doing relative to everyone else?' is a better design goal for some contexts. Designers might therefore wish to *avoid* supporting fairness in some situations. This may seem like an odd claim given the subject matter of this thesis, but we believe that identifying situations where supporting fairness is detrimental to group work is just as important as identifying those situations where it is critical. Furthermore, information about contributions might not even be useful in some work processes. Using our earlier example of collaborative writing, is it really helpful to see all the minor alterations that have been made in a document? Does knowing that someone spent an hour writing a paragraph only to delete it shortly after really encourage fairness? Clearly this is a design space that warrants further research.

In addition to these questions, there are broader issues that are raised by the subject of designing for fairness. For example, have suggested that designers have the choice between hiding or sharing information that could help or hinder group activity. But should a designer engage in such activity? Is it right or wrong to make such choices about what people can or cannot see? We suggest that there are important but as yet unexposed ethical consequences, both for designers and the people to whom their designs cater, that should be mapped out in future research.

8.2.3.1 Design Guidelines for Supporting Fairness

Based on the discussion above and the overall findings of this thesis, we offer the following points for designers to consider when thinking about fairness in collaborative systems.

- The work process involves multiple stages, and fairness plays a different role in each. During the assignment of work (planning phase), collaborators need to agree on distributions of workload. Designers responsible for implementing division of labour should try to account for basic notions of equality but must allow policies to be adjusted in light of group preferences. An example would be division by competencies versus division by equality.
- However, when thinking about division of labour, designers should be mindful of the fact that fairness is malleable, and that different rules and norms are followed depending on the work situation, the relationships between individuals, and the type of work that is being done. It is therefore not possible to dictate in advance what is and is not fair when implementing policies that guide the assignment of work. Designers should instead aim to provide people with information that can be used to make such judgements, rather than attempt to prescribe fairness in advance.
- In the process of completing work (enactment phase) designers should aim to allow people to check whether collaborators are adhering to commitments. In order to achieve this, collaborators need information about the work that is being done. Such information can be more or less obscure, but fairness judgements are likely to be most accurate when contributions are described at a reasonable level of detail. In addition, fairness is a social judgement, and in order to evaluate fairness in a work process, designers should allow people to make comparisons between contributions.
- Although supporting fairness can be helpful, a question exists as to whether one should design for fairness at all. There may be situations in which support for fairness is not required, such as those where task completion is not the primary focus. Some work settings might actually be hindered by support that encourages people to reflect upon fairness, especially if reflection leads to dissension or altercation. Designers should be aware of the limitations of prior examples (e.g. meters in Warcraft) and consider the potential consequences of making the actions of others public.

8.3 A Theoretical Model of Division of Labour and Fairness Judgements

In this section we provide a theoretical contribution by offering a model of the process by which individuals establish and iteratively assess fairness during collaborative work. The model seeks to expose the decision process an individual would follow when trying to determine an appropriate allocation of work and, later, the process by which fairness would be evaluated during the enactment of agreed allocations. The model helps to formalise and draw together our most recent thinking on the subject of how people make judgements about fairness during collaboration. The model has its roots in theories of collaborative work and models of distributional justice judgements. In creating this new model we have drawn on our thesis work to consider how particular findings serve as evidence about particular decision outcomes, and how our acquired knowledge of fairness can provide a base to understand decisions in the context of collaboration. The following subsections present the rationale for our model and the decision processes it aims to capture. Finally, we consider the various stages of the model with respect to our thesis findings.

8.3.1 Foundations

Our model is instatiated by first drawing on the conceptualisation of collaborative work offered by Galegher & Kraut (1990). Their outline (previously reviewed in Chapter 2 of this thesis) represents a theoretical position about the different stages involved in collaboration. We use their outline as a starting point because it draws a clear delineation between planning and enactment of work, allowing us to consider how fairness might be assessed at each stage. As a brief review, Galegher & Kraut view collaboration as formed of three stages:

1. Planning, which encapsulates initial division of labour.
2. Enactment, where agreed allocations are carried out.
3. Integration, in which individual work packages are interwoven into a coherent whole.

Our model uses the stages of planning and enacting as platforms for a larger framework that represents the decision processes an individual might follow when assessing fairness during the division and completion of collaborative work. It is worth reiterating that these stages may be neither linear nor fixed; some collaborations might proceed as outlined but others might require transitions where initial workloads need to be readjusted during enactment. The extent to which task components are coupled will also affect conformity to this process—it might be possible to contribute loosely coupled bits of work on-the-fly (as when contributing items to a collaborative bibliography), whereas other work packages might require careful integration, as might occur in collaborative software development. Our proposed model can account for this by providing scope for decisions about distributive and enacted fairness to be taken several times prior to the end of a collaborative process.

In describing this model, our primary aim is to account for the various cognitive checks an individual would need to perform in order to determine whether the distribution and completion of work is in line with perceptions about fairness. Our considerations of distributive fairness in the planning phase of work are inspired by Gerald Leventhal's justice judgement model (Leventhal, 1976, 1980). His model articulates the sequence of steps an individual would follow when trying to evaluate distributive fairness. Our contribution is one of adapting the decision sequence such that it is directly oriented towards decisions about fairness in the division of workload. Our conceptualisation of stage 2, Enactment, provides an entirely novel decision procedure that captures the evaluation of fair behaviour in an ongoing work process. Fair behaviour in this case is defined in terms of the extent to which an individual can be said to be *adhering to their commitments*, as established by the initial division of labour.

The proposed model is described as follows:

1. Planning: Assessment of fairness in division of labour.

   1.1. Rule collection and weighting.

   1.2. Preliminary estimation.

   1.3. Rule combination.

   1.4. Outcome evaluation.

2. Enactment: Assessment of fairness in completion of work.

   2.1. Identification of commitments and rules.

   2.2. Comparison of commitments to task progress.

   2.3. Seek explanatory rationale.

   2.4. Outcome evaluation.

Because we did not study integration in this thesis, we place it aside and consider it as a task for future work. In what follows, we elaborate on the stages listed above detail and outline assumptions that support particular phases of judgement.

8.3.2 Assessing Fairness in Allocation

The planning stage of work involves decisions about division of labour and encapsulates decisions about *who* is going to do *how much* of *what*, as well as when, where, and how work will be done. Fairness at this stage is therefore relevant to the assignment of work in line with relevant norms, as well as consideration of rationale that would provide deviation from said norms. As with our earlier definition of fairness, our consideration of how individuals assess fairness in the allocation of work is underpinned by the basic assumption of *equality*. That is, if no rationale exist to suggest otherwise, the expectation is that everyone will contribute an equal (or at least very similar) amount of work. From this default position, the following judgement sequence (cf. Leventhal, 1980) is hypothesised to occur when making an allocation:

1. *Rule collection and weighting*: an individual decides which distribution rules are appropriate and determines the relative importance of said rules. More important rules are assigned a higher weighting, where weight corresponds to a greater impact on the perception of fairness (Leventhal, 1980). An example of two rules relevant to division of labour would be *divide equally* versus *divide by competencies*.

2. *Preliminary estimation*: the individual estimates the work that should be assigned based on the rules that have been determined as relevant to the current situation. This might involve estimations about the amount and type of work to be assigned. For example, if the equality rule has been invoked, the individual performs an estimation that creates an equal distribution of work. If competencies are considered relevant, the individual might estimate which work packages fit best to the skillsets of particular individuals.

3. *Rule combination*: the individual combines preliminary estimates with considerations of other factors to arrive at a final judgement about what is fair. In this case, 'other factors' would be those that provide *legitimate* rationale for deviations from equality. Such factors go beyond deviations implied by particular assignment rules and incorporate knowledge about issues including reputation, perceived entitlements, and mitigating circumstances. For example, if work has been assigned on the basis of equality, but one individual has been enlisted to work on several other projects simultaneously, it might be regarded as fairer to lower the person's workload to restore parity. However, this might not be acceptable if there are other factors relevant to the situation that would make such a shift intolerable. In addition, this stage might involve considerations about *efficiency*—whether inequitable workloads might provide a balance of effort in relation to time. Thus this stage is fundamentally about attending to extraneous issues that would give cause for legitimate deviations from equality.

4. *Outcome evaluation*: In the final stage of the sequence, the individual assesses the fairness of each person's allocation. This involves comparison between estimates and reality in order to assess whether further iteration on the allocation procedure is required.

If the allocation is considered fair, work can proceed as assigned. However, if the individual perceives unfairness in the final division of labour, he or she might suggest a different allocation by invoking new rules, or by adjusting the weights of current rules, to revisit the estimation and combination stages until a satisfactory outcome is reached. Thus, the decision model has sufficient scope to allow for cycles through the various stages. In terms of our prior work, we can think about these considerations in terms of an iterative DLUG—an individual might suggest an allocation, see it rejected, and then propose a second assignment based on rules relevant to the situation. It should be noted that we regard the overall decision process as largely implicit, in that it is not something that the individual will necessarily be active in cognizing. Decisions such as these are likely to be well practiced and hence largely unnoticeable to the individual (Leventhal, 1980).

8.3.3 Assessing Fairness in Enactment

The second stage of the model considers how an individual would assess fairness during the work process. We introduce this stage with two initial arguments. First, we suggest that this stage is fundamentally about *monitoring commitments*, where each person's commitment is defined as the work that they have each agreed to enact as a result of the most recent division of labour. Thus, tracking fairness at this stage involves a process of checking whether people are *adhering* to their commitments, and this occurs through natural monitoring of the actions of others. Our second argument emphasises the idea of monitoring as critical because, based on the literature on collaborative work, it is well known that people actively seek information to create a sense of awareness during collaboration (Gutwin & Greenberg, 2002, see Chapter 7 of this thesis). Gathering such information not only allows one to understand the actions of collaborators but also provides a context for one's own actions (Dourish & Bellotti, 1992). We suggest that, in collecting awareness information, people naturally make inferences about the work that is being done and, through making those inferences, may gain sufficient cause to question whether or not the current state of affairs is in line with agreed allocations. Thus we do not argue that people actively 'monitor fairness' but rather that the collection of awareness information may give cause to question whether each person is adhering to their commitments, e.g. through cues that suggest something is wrong. This distinction feels important as we must recognise that fairness is only one motivation among many during collaboration. We suggest that fairness is a concern that, once established, remains in the background most of the time—if everything is fair, everything is good, and there is no need to monitor it—but situations may arise that bring the experience of *unfairness* to the fore. Thus we regard fairness as similar to coordination in that it tends to be most evident when it is lacking, and it is at those times that individuals will have reason to make decisions in an attempt to re-establish harmonious working.

Since no research has previously considered fairness in the enactment of work, the following scenario is a novel outline of the process an individual might follow when attempting to resolve dissonance between current behaviour and perceptions about fairness. Based on our arguments about adhering to commitments, as well the present thesis findings, we suggest that an individual follows these four steps:

1. *Identification of commitments and rules*: the individual draws on his or her knowledge of the collaboration to identify each person's commitments and the rules that were followed to arrive at the initial division of labour. This represents an attempt at forming an understanding of what the work process *should* look like. It may be that in some situations, no explicit allocation was agreed—the implicit commitment in this setting can be regarded as the defaul of an equal contribution.

2. *Comparison of commitments to task progress*: this stage involves comparing each person's commitments to the current status of work in an attempt to identify disjoints between the two. Perceived disjoints would indicate deviation from commitments. This stage might involve comparisons between quantitative workloads, but an individual might also form suppositions about the effort that is being invested into the task, as measured in terms of either quality or workrate.

3. *Seek explanatory rationale*: the individual looks for reasons that would explain or legitimise deviations from commitments. This would involve drawing on knowledge about one's collaborators. For example, if it is known that an individual has been unwell and thus was unable to contribute, policing of fairness might be relaxed. This step has sufficient scope to capture a wide range of scenarios; for example, rationale might be derived from distributions of power, control, or ownership of task components.

4. *Outcome evaluation*: In the final stage of the sequence, the individual determines whether the current state of affairs is acceptable with respect to his or her current understanding of fairness. This is essentially a test of whether deviations from commitments are tolerable or whether action needs to be taken to change the current state of affairs and enforce fairness.

If the outcome of the fourth step is positive, the individual can continue working, perhaps with an updated understanding of what is fair in the current situation. There might be some adjustment of behaviour, e.g. an increased level of effort, or reassignment of tasks to cope with what is happening. It may be that mitigating circumstances have affected an individual's ability to participate in the work process, and thus some of his or her work should be redistributed. On the other hand, contextual factors might make deviations from equality highly tolerable.

However, if the outcome of the decision is negative, the individual must take further action on the person or persons to reinstatiate fairness. The chosen action might be taken from the following options:

- Confront individual. (Make the other person aware of and accountable for their unfairness in an attempt at repairing the collaboration).

- Defer payment. (Recognise and tolerate the act of unfairness, continue collaboration, but invoke the act later down the line if unfairness persists).

- Impose sanction. (Penalise the individual by agreeing to withdraw certain rewards, but allow the collaboration to continue).

- Reciprocate. (Do something similar to the other person, such as exclude them from project meetings).

- Renege on collaboration altogether. (Defect on current situation and avoid working with the person in future).

These actions range from attempts at conciliatory behaviour to ending the collaboration entirely. It is likely that the individual will explore less extreme options before choosing to defect. Thus, as with the process followed during the planning phase, an individual might engage in multiple iterations through the decision path before choosing to defect. We also suggest that each of the options might be followed by the reassignment of work—an initial confrontation might be enough to determine that workloads must be reassigned. What is useful is that the model has sufficient scope to allow for *self-confrontation*. That is, if an individual feels that they themselves are not conforming with fairness norms, they may seek repair by offering to take work from their colleagues.

This model is, of course, a first attempt at building a theory of how fairness plays out in collaborative division of labour. There may be issues that are not well-handled by the framework we have proposed, and thus a challenge for future work will be to identify such issues and refine the model until it can account for the available evidence.

8.3.3.1 Linking the Present Thesis Findings to the Model

Here we consider how various findings from this thesis relate to different parts of the model. Considering each finding relative to the model helps to provide further justification for the arguments we have presented.

- In thesis study 1 we found that student workgroups desired fairness in the allocation and completion of collaborative tasks. Their basic *expectation of equality* underpins the model, and their use of other allocation procedures (e.g. assignment by competency) is reflected in step 1.1 of the planning phase, where individuals are assumed to invoke relevant rules to guide the initial assignment of workloads.

- In the DLUG studies, participants preferred equal allocations of work, even in circumstances where they could not be punished for unfairness. This further emphasizes the basic premise of equality in division of labour. In addition, the fact that several experimental manipulations did not give cause to sway from the equal allocation suggests that the impact of the variables was not enough to provide *sufficient rationale* for deviation from equality. We assume that individuals traversed step 1.3 of the planning phase model and failed to identify (or simply chose to ignore) other factors that could have legitimised deviations from equality.

- During completion of the work assigned via the DLUG, some collaborators reworked initial assignments during the process of work. This may be evidence of *comparing commitments to task progress* to determine that the assignment of work should be renegotiated (step 2.2 of the enactment phase). The fact that the majority of pairs did the work as agreed is evidence of *adhering to commitments* as a motivating force when completing assigned workloads.

- The yoking of completion times discovered in the DLUG experiments may be evidence of *comparisons about fairness in workrate*. That is, cues about the partner's efforts may have given sufficient cause to ask questions about whether the current investment was appropriate for the task at hand. In some cases, this may have lead individuals to speed up or slow down until an 'equitable division of labour' was achieved (Jackson & Harkins, 1985). The implication here is that a quantatively equal allocation might also give rise to a norm of equality in the process itself. The checking of such norms would be encapsulated by steps 2.1 and 2.2 of the enactment phase.

- We found in study 6 that individuals were highly tolerant of unfairness in casual work settings. This may be the result of having *collected rationale* to suggest that the work partner, current context, and task demands were sufficient to allow deviations from equality to be tolerated. This is captured by step 2.3 of the model for the enactment phase.

- We suggested in study 7 that designers could support fairness by allowing individuals to monitor and compare the contributions of their collaborators. This is in line with the foundations of the decision model at the enactment stage, where we suggest that observation of contributions can give rise to questions about fairness. Moreover, the statements offered by players in World of Warcraft suggest that information about individual contributions is collected as a means of checking adherence to commitments, and gives cause to seek rationale (2.3 of the enactment phase) to explain deviations from the required standard. Sanctions that are imposed upon players (e.g. temporary exclusion from future team efforts) would arise as a result of following the decision model and opting to *take action* as a result of perceived unfairness.

The model we have proposed may help future scholars to consider the steps employed to evaluate fairness, but might also be used to implement policies intended for arbitration in collaborative systems. For example, online marketplaces like *eBay* require arbitration procedures to resolve disputes between two or more parties. Our decision model could help designers to think through the steps by which fairness is evaluated, such that systems can be better designed for fairness. For example, the designer might consider increasing the salience of particular rationale that would help explain to a buyer while a seller appears to be deviating from commitments. In situations more relevant to collaborative work, the decision model could be employed to capture certain types of information that would support the process of fair judgement. For example, in the knowledge that people are assumed to expect equality and seek rationale to explain or legitimise inequality, designers could think about showing collaborative contributions in a timeline view such that an individual's ability to evaluate holistic contributions is strengthened. Allowing annotation of lapses in activity might allow individuals to seek rationale from their coworkers, in turn providing cause to adjust or strengthen their own understanding of fairness and the current work process.

8.4 Thesis Limitations and Future Work

At the beginning of this thesis (Section 1.2) we described some initial delimitations that were sufficient to scope the research reported here. For example, while our work on division of labour raised fairness as an issue, we elected not to consider the issue of retributive fairness. Some study-specific limitations have also been described within the relevant chapters. This section discusses overall limitations that became apparent during the progress of this research. We regard these limitations as opportunities for future work.

8.4.1 Participants and Generalizability

One of the main limitations of the research reported in this thesis is the relatively small sample size for each study. Additionally, all of the participants recruited for our studies were self-selecting, which is a further limitation given that the behavioural tendencies of self-selecting volunteers may not be representative of the broader population (cf. Rosenthal & Rosnow, 1969). Finally, many of our participants were undergraduate students, and there is some contention within the literature as to how well the preferences of students match those of others in the outside world (e.g. Eckel & Grossman, 2000; Fehr & Fischbacher, 2004; Henrich *et al*., 2010; Engel, 2011). These issues mean that we cannot assume that our results are representative of other populations.

While we do not believe that any of our conclusions are compromised by the low number of participants, it is possible that some manipulations (e.g. those using the DLUG dictator game, as in Chapter 5) would have statistically reliable effects in much larger experiments. While the problems associated with sample size naturally limit our ability to apply our findings to other real-world settings, we do not believe that they are enough to invalidate the findings of this work. We have been careful to avoid overreaching the generalisability of our findings, and, where appropriate, we have emphasized that particular findings are illustrative rather than representative. Beyond relying on co-located human participants, one alternative approach would have been to use crowdsourcing platforms like Amazon's 'Mechanical Turk'[38] to reach a large number of participants from afar. However, Mechanical Turk was not available in the UK at the time our studies were conducted. Several other platforms have since become available, making a crowdsourced exploration of the DLUG a viable opportunity for future work.

8.4.2 Fairness in Work: Definitions and Cultural Sensitivities

In drawing attention to the role of fairness and equality in division of labour, our studies necessarily entailed certain definitions of terms. Our conception of fairness is based on the idea that each person should be attributed his or her dues in line with perceived entitlements, rights, or needs. Further, the norm of equality was used as a basis for our understanding of fairness—where no entitlements, rights, or specific needs exist, fairness 'defaults' to the expectation of an equal distribution. In line with this, people use

[38]Mechanical Turk, see: https://www.mturk.com/mturk/welcome

equality as a decision-making heuristic during the allocation of goods and services (Allison & Messick, 1990).

However, our definitions, as well as the research supporting them, are based on a primarily Western school of thought. Based on our results, our definition appears to reflect the behaviour of our participants, most of whom were interested in fairness and expected others to behave in line with equality. Our findings may be reflective of Western norms but are unlikely to apply to other cultures because what is regarded as a fair offer in one culture is sometimes perceived quite differently in another. For example, in Chapter 4, we described how the majority of studies on the classic UG reveal a prevalence of equality over selfishness. This equality can be regarded as compliance with fairness norms. Yet studies of non-industrialised and pre-technological socities demonstrate differences in social constructions of 'fairness' (Henrich *et al*., 2001). In some societies, people are not inclined to make equitable offers and do not punish those who make low offers, yet in others, people were extremely generous and frequently offered more than half of the prize to responders (Henrich *et al*., 2001).

Since expectations about what is fair are not universally consistent, it seems clear that expectations about fairness in division of labour would also differ between contexts. However, we do not believe these concerns impact our basic definition; rather, cultural differences mean that expectations about rights and entitlements are likely to be different, further emphasizing the very subjective nature of fairness. What they may impact, however, is the basic expectation of equality in allocations. Cultural differences might also impact the extent to which people are perturbed by deviations from equity in their work process. As a challenge that was well beyond the reach of this thesis, future work should explore the extent to which our findings are relevant to other societies.

8.4.3 Factors Affecting Division of Labour

While we did explore some factors that could influence fairness in division of labour, e.g. the potential threat of punishment, the present work has not been able to account for an abundance of other factors that will impact division of work in the real world. Such factors would be prevalent in everyday workplace settings. Examples include, and are certainly not limited to, organisational protocols, power differentials and status hierarchies (Eason, 1996; Bardram, 1997); perceptions about gender roles (Mikula, 1998); job roles that prescribe assignment of tasks (Symon *et al*., 1996); skills, knowledge, and individual expertise (Grinter *et al*., 1999); pecuniary incentives (Camerer & Hogarth, 1999); and standard operating procedures (Grinter, 1996). Such issues are bound to influence the extent to which people perceive a need to conform with equality during collaborative work.

Of course, such factors are too numerous to explore within the confines of a single thesis. We suggest that they are instead worthy of exploration in further work, which could be achieved using our DLUG. Perhaps a more important issue is whether or not the above factors affected our empirical results. On the whole, we believe this to be unlikely because most of the groups in our studies were egalitarian in their distribution of power. This is at least true of the student workgroups in Chapter 3 and the pairs in Chapter

6. It is also true of the anonymous pairings in Chapter 5, although one might argue that assignment of roles in the DLUG does imply a certain status difference. Certain elements of this work might have been affected by knowledge and skills, but any such effect appears minimal given the prevalence of equality in offer data.

8.4.3.1 Potential Elaborations of the DLUG

Below we offer some specific ideas as to how DLUG model proposed in Chapter 4 may help illuminate a range of issues relevant to economic decision making and division of labour. Our list is necessarily selective, and we eschew more esoteric manipulations in favour of those that pertain either directly to the results of this thesis or to the potential for understanding collaborative work practices.

Iterate the allocation phase of the DLUG. An iterated version of the DLUG might be adopted, where a rejected offer allows the proposer to revise their offer or end the game (cf. Gneezy *et al*., 2003). Such iteration would also bring the model closer to a genuine negotiation process, and might allow for the investigation of strategy selection (Brenner & Vriend, 2006; List & Cherry, 2000) or deadlines (Gneezy *et al*., 2003).

Repeat resolved DLUGs. Experiments where players perform multiple collaborative work tasks would allow richer investigation of the role of social relations in division of labour. Tying in with our earlier discussions on competence and expertise, participants might decide maximum efficiency can be achieved by dividing work differently after a single round of obtaining five items apiece. In any case, it would be interesting to explore whether people learn to become more self-interested in division of work, given an absence of other incentives to cooperate.

Impose power differentials. Allocating the roles of proposer and responder according to causal factors, e.g., relative performance on a prior task (Hoffman *et al*., 1994), would encourage participants to interpret the roles of proposer and responder as signalling differences in status. This would play further on perceptions of entitlement, which could potentially cause participants to shift allocations in line with fairness norms. Studies of the UG have shown that manipulations of this kind allow explorations of resource entitlements (Hoffman *et al*., 1994; List & Cherry, 2000), and such designs might be leveraged to model status hierarchies in division of labour.

Increase the interdependency of work items. In the present work we focused on collaborative search tasks. Collaborative search can be characterised as relatively loosely coupled work, in the sense that none of the work items to be distributed is especially dependent on any of the others. Increasing interdependency by using different tasks, e.g., collaborative writing or programming, would allow for comparisons between different work contexts. It might also lead to different observations of coordinative activity, in the same way that we observed differential strategy use during collaborative information seeking.

At a broader level, our DLUG could be used in studies of agent-based negotiation. A common approach in social science research is to substitute human participants with simulations where agents are paired off and must negotiate or complete some task. Economic games are used in many such experiments (e.g. Brenner & Vriend, 2006; de Jong *et al*., 2008). The DLUG could easily slot into these settings and be used for novel ends beyond those reported in this thesis.

8.4.4 Task Choices, Work Settings, and Group Construction

Within this thesis, the scope of studies was limited to a few very specific types of work. Our DLUG experiments focused on collaborative information seeking, which we then carried over into the field study reported in Chapter 6. While this choice was made for purposes of consistency, and because division of labour is a relevant problem for collaborative search, our findings are limited by this choice. Future work must explore division of labour in other collaborative tasks.

A second limitation in this area is that, although there is some overlap within certain studies, the majority of work in this thesis has focused on synchronous collaboration. Future work could examine how our findings map to asynchronous work. For example, the matching effect discovered in Chapter 5 might be more or less prevalent in asynchronous collaboration—if collaborators have information about how long their colleagues spent on a task, they might use this information as a rough guide to manage their own working period and investment of effort. On the other hand, collaborators might be more interested in catching up with their coworkers' effors rather than striving for equality. If the task structure was kept the same as in the experiments we reported, asynchronous work could have a radical effect on coordination problems and strategies (e.g. all the responsibility for non-duplication might fall on later work). The design ideas and findings presented in Chapter 7 could also be extended to asynchronous interaction. Although awareness metrics for judging contributions might also be useful in asynchronous work, perhaps the time lag might increase the need for detail about certain prior efforts over others.

Lastly, findings from the empirical studies in Chapters 5 and 6 were isolated to dyadic interaction. Managing fairness is likely to be more difficult as a group becomes larger; in particular, diffusion of responsibility implies a greater risk of loafing behaviour (e.g. Latane *et al*., 1979; Jackson & Harkins, 1985; Kravitz & Martin, 1986) and this is likely to be even more problematic in computer-mediated settings (Kraut, 2003) where identifiability is low and task progress is difficult to ascertain. Our focus on dyads was partly to minimise difficulties with participant recruitment, but was perhaps also influenced by our decision to adapt the classic UG, a paradigm that typically involves only two persons. Initial evidence from our survey study in Chapter 3 implies that fairness remains relevant in larger groups. Future work should aim to extend our empirical work to groups of three or more.

8.4.5 Research Methods

The studies undertaken in pursuit of this thesis used a variety of approaches, each of which has generated interesting findings relevant to our topics of concern. Each of these approaches does, however, have its limitations. Here we discuss some foremost limitations associated with our choice of methods.

In our explorations of the DLUG, we did not have the opportunity to explore every possible idea, and thus we aimed for a stepwise approach in the vein of incremental experimentation. While this approach is ideal for reductionist empirical science, it may be viewed with suspicion by certain researchers in CSCW. We hope that our findings, especially the matching effect in search times, are enough to suggest that the DLUG can play a role in developing our understanding of CSCW more broadly.

There are two more general limitations associated with our experimental approach. One is the use of laboratory methods, which necessarily opens oneself up to the sorts of criticisms aimed towards any lab-based research. Within HCI in particular, there is an ongoing discourse about the utility of laboratory methods and their relevance to 'in-the-wild' behaviour (e.g. Marshall *et al.*, 2011; Rogers, 2011; Hornecker & Nicol, 2012). Such accounts argue that laboratory findings are not sufficiently generalisable, and that experimental setups are typically too rarefied to understand how technologies are used in the extra-lab world. Based on the latter premise, Rogers (2011) calls for 'in-the-wild theory development' based on *in-situ* studies that account for social context and the less controllable aspects of human behaviour. This aligns with broader trends in CSCW, where ethnographic workplace studies are often championed because they produce detailed accounts of organisational behaviour in the real world. However, what researchers often fail to recognise is that workplace and other field studies are themselves highly contextualised, which in turn makes their results difficult to generalise (cf. Kraut, 2003). Moreover, disjoints between theory and empirical observations usually imply a need for more, not fewer, laboratory studies; it is only then that theory and data can be reconciled, as is ongoing within the literature on economic rationality and social preferences from which our DLUG was derived. Nevertheless, our reductionist approach would be well-complemented by further and more holistic studies of collaborative work. We note that, with the studies reported in Chapters 6 and 7, this thesis did take some initial steps in this regard, and we hope our work was made stronger through combining quantitative and qualitative approaches.

A second limitation is related to the very nature of economic models. In Chapter 4 we showed that economic games have seen prior use in HCI. One benefit of these models is their simplicity. Each allows for the isolation of individual variables, which can then be systematically manipulated to explore their effect on fairness (or cooperation, as the outcome is usually operationalised within HCI research). The downside of these models is their scope—they record a single decision, taken out of context at a particular moment in time. Our DLUG paradigm offers a small iteration on this limitation; not only did we explore the planning phase of work using our model, we then studied the enactment of the shared task after a quantitative allocation had been reached. Yet, as an abstracted version of workload allocation, our model still neglects issues that would prevalent in the real-world. For instance, it is likely that collaborators would engage in multiple rounds of workload allocation, contributing and evaluating suggestions until a suitable allocation is achieved (Freidson, 1976; Strauss, 1985). We note that our DLUG can be expanded to model these aspects of work assignment. Additionally, work items in the real world will not be so easily packaged into numeric form, as in our DLUG studies. Tasks are often fuzzy and may contain interdependent elements that cannot be easily decoupled (Herbsleb & Grinter, 1999). Our experiments had to simplify these aspects of work, but perhaps future studies of the DLUG could employ work items that are of variable difficulty, more or less interdependent, or have other features that we did not test here.

The studies reported in Chapters 6 and 7 used qualitative techniques to analyse interview data. One limitation of qualitative analysis is that data interpretation is very much dependent on the view of the analyst—this is not a fundamental flaw but does mean that other interpretations are possible (Sandelowski, 1995; Elo & Kyngas, 2007). However, there is no way of checking which interpretation would be more meaningful. Both of our studies used unguided, inductive coding. We felt that this approach was appropriate given that, while we had some broad topics of concern, we were not testing any theories nor seeking to answer any specific questions about behaviour within the two studies (Elo & Kyngas, 2007). A downside of our approach is that we cannot make claims about prevalence, nor importance, as would be possible in a study where deductive coding and inter-rater reliabilities were utilised. Finally, our reliance on interview data means that both studies were affected by the very nature of retrospective self-reports (cf. Nisbett & Wilson, 1977). For instance, our results are reliant upon what our participants were able to recall about their experiences, and might also have been tempered by our choice of interview questions.

8.5 Concluding Remarks

This thesis has investigated the issues of fairness and division of labour in collaborative work situations. Our findings suggest that the distribution and completion of work is guided by concerns about fairness: collaborators allocate work so as to reach notionally fair assignments, and then expect those assignments to be completed in line with fairness norms. However, ensuring fairness can be difficult under circumstances of limited awareness, as is often the case in computer-mediated settings. We argued that assessment of fairness is based on the judged legitimacy of deviations from equality, and introduced a design solution that could support fairness by providing collaborators in a CSCW system with information about the contributions provided by each team member. We argued that such a mechanism should allow for comparison of contributions relative to one another, and should provide sufficient detail about said contributions so as to make social comparisons meaningful.

This thesis has advanced the field by specifying the relevance of fairness to division of labour during collaborative work, as well as to CSCW more generally. Our efforts have united several previously disparate literatures, and our investigations have resulted in a number of pragmatic, theoretical, and empirical contributions. These contributions are not solely related to division of labour but also touch on other research areas, including collaborative information seeking and collaborative gaming within virtual worlds. Additionally, we presented an economic approach to studying fairness in the division of labour, offering a methodological contribution that is novel to scientific inquiry. Yet despite these contributions, we regard the work of this thesis as only the first step towards a more complete understanding of how CSCW systems could, and perhaps should, support fairness.

References

Akerlof, G., & Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics*, *115*(3), 715–753.

Alge, B. J. (2001). Effects of computer surveillance on perceptions of privacy and procedural justice. *Journal of Applied Psychology*, *86*(4), 797–804.

Allison, S. T., & Messick, D. M. (1990). Social decision heuristics in the use of shared resources. *Journal of Behavioral Decision Making*, *3*, 195–204.

Amershi, S., & Morris, M. R. (2008). Cosearch: A system for co-located collaborative web search. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, (pp. 1647–1656). New York, NY, USA: ACM.

Andersen, S., Ertac, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, *101*, 3427–3439.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Ariely, D. (2008). *Predictably Irrational*. London, UK: HarperCollins.

Aula, A., & Haki, M. (2008). Less is more in web search interfaces for older adults. *First Monday*, *10*(4).

Bailey, E. W., Kelly, D., & Gyllstrom, K. (2009). Undergraduates' evaluations of assigned search topics. In: *Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval*, SIGIR '09, (pp. 812–813). New York, NY, USA: ACM.

Bannon, L., & Schmidt, K. (1991). CSCW: four characters in search of a context. In: J. M. Bowers, & S. D. Benford (Eds.) *Studies in Computer Supported Cooperative Work*, (pp. 3–16). Amsterdam: North Holland: Elsevier Science Publishers.

Bardram, J. E. (1997). Plans as situated action: an activity theory approach to workflow systems. In: *Proceedings of the fifth conference on European Conference on Computer-Supported Cooperative Work*, (pp. 17–32). Norwell, MA, USA: Kluwer Academic Publishers.

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, *11*, 122–133.

Bardzell, J., Nichols, J., Pace, T., & Bardzell, S. (2012). Come meet me at ulduar: progression raiding in world of warcraft. In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, (pp. 603–612). New York, NY, USA: ACM.

Bardzell, S. (2010). Feminist hci: taking stock and outlining an agenda for design. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1301–1310). New York, NY, USA: ACM.

Bardzell, S., Bardzell, J., Pace, T., & Reed, K. (2008). Blissfully productive: grouping and cooperation in world of warcraft instance runs. In: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, (pp. 357–360). New York, NY, USA: ACM.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. New York, NY: Wiley.

Bayerl, P. S., & Lauche, K. (2008). Coordinating high-interdependency tasks in asymmetric distributed teams. In: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, (pp. 417–426). New York, NY, USA: ACM.

Ben-Ner, A., Kong, F., & Putterman, L. (2004). Share and share alike? gender pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, *25*, 581–589.

Benkler, Y. (2012). The penguin and the leviathan: Towards cooperative human systems design. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, (pp. 1–2). New York, NY, USA: ACM.

Benton, L., Johnson, H., Ashwin, E., Brosnan, M., & Grawemeyer, B. (2012). Developing ideas: supporting children with autism within a participatory design team. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 2599–2608). New York, NY, USA: ACM.

Berger, R., Rauhut, H., Prade, S., & Helbing, D. (2011). Bargaining over waiting time in ultimatum game experiments. *Social Science Research*, *41*, 372–379.

Binmore, K. (2010a). Interpersonal comparison in egalitarian societies. *European Journal of Political Economy*, *26*(3), 294–301.

Binmore, K. (2010b). Social norms or social preferences? *Mind & Society*, *9*(2), 139–157.

Binmore, K., Shaked, A., & Sutton, J. (1985). Testing noncooperative bargaining theory: A preliminary study. *American Economic Review*, *75*, 1178–1180.

Birnbaum, I. (2013). World of warcraft lost more than a million subscribers last quarter, is still bigger than everything ever. PC Gamer: Available Online: http://www.cnet.com.au/world-of-warcraft-loses-1-3m-players-339344234.htm.

Bly, S. A., Harrison, S. R., & Irwin, S. (1993). Media spaces: bringing people together in a video, audio, and computing environment. *Communications of the ACM*, *36*(1), 28–46.

Boellstorf, T. (2008). *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. USA: Princeton University Press.

Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, *89*(1), 335–339.

Bolton, G. E. (1991). A comparative model of bargaining behaviour. *The American Economic Review*, *81*, 1096–1136.

Bolton, G. E., Katock, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, *27*, 269–299.

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior*, *10*(1), 95–121.

Borning, A., & Muller, M. (2012). Next steps for value sensitive design. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1125–1134). New York, NY, USA: ACM.

Bos, N., Olson, J., Gergle, D., Olson, G., & Wright, Z. (2002). Effects of four computer-mediated communications channels on trust development. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '02, (pp. 135–140). New York, NY, USA: ACM.

Bos, N., Olson, J., Nan, N., Shami, N. S., Hoch, S., & Johnston, E. (2006). Collocation bindness in partially distributed groups: is there a downside to being collocated? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, (pp. 1313–1321). New York, NY, USA: ACM.

Bradner, E., & Mark, G. (2002). Why distance matters: effects on cooperation, persuasion and deception. In: *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, CSCW '02, (pp. 226–235). New York, NY, USA: ACM.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brenner, T., & Vriend, N. J. (2006). On the behaviour of proposers in ultimatum games. *Journal of Economic Behavior & Organization*, *61*, 617–631.

Brooks, C. M., & Ammons, J. L. (2003). Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments. *Journal of Education for Business*, *78*, 268–272.

Brown, B., & Baarkhuus, L. (2007). Leisure and CSCW: introduction to special issue. *Computer Supported Cooperative Work*, *16*, 1–10.

Brown, B., & Bell, M. (2004). CSCW at play: 'There' as a collaborative virtual environment. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, (pp. 350–359). New York, NY, USA: ACM.

Buchan, N. R., Croson, R. T. A., & Johnson, E. J. (2004). When do fair beliefs influence bargaining? experimental bargaining in japan and the united states. *The Journal of Consumer Research*, *31*(1), 181–190.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*(3), 1014–1027.

Camerer, C., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*, 7–42.

Camerer, C. F. (2003). *Behavioral Game Theory*. USA: Princeton University Press.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: ultimatums, dictators and manners. *Journal of Economic Perspectives*, *9*(2), 209–219.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, *37*(1), 47–59.

Capra, R. (2010). Hci browser: A tool for studying web search behavior. In: *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem, Volume 47*, ASIS&T '10, (pp. 175:1–175:2). Silver Springs, MD, USA: American Society for Information Science.

Capra, R., Chen, A. T., Hawthorne, K., Arguello, J., Shaw, L., & Marchionini, G. (2012). Design and evaluation of a system to support collaborative search. In: *Proc. ASIST*.

Capra, R., Marchionini, G., Velasco-Martin, J., & Muller, K. (2010). Tools-at-hand and learning in multi-session, collaborative search. In: *Proc. CHI*, (pp. 951–960).

Carlsson, F., He, H., & Martinsson, P. (2013). Easy come, easy go: The role of windfall money in lab and field experiments. *Experimental Economics*, *16*, 190–207.

Carter, J., & Irons, M. (1991). Are economists different, and if so why? *Journal of Economic Perspectives*, *5*, 171–177.

Cason, T. N., & Mui, V. L. (1998). Social influence and the strategy method in the sequential dictator game. *Journal of Mathematical Psychology*, *42*, 248–265.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), 47–83.

Chen, K., & Tang, F. (2009). Cultural differences between tibetans and ethnic han chinese in ultimatum bargaining experiments. *European Journal of Political Economy*, *25*, 78–84.

Chen, M. (2008). Communication, coordination, and camaraderie in world of warcraft. *Games and Culture*, *4*(1), 47–73.

Chen, Y., & Gazzale, R. (2004). When does learning in games generate convergence to Nash equilibria? The role of supermodularity in an experimental setting. *American Economic Review*, *94*(5), 1505–1535.

Cherry, T. L. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology*, *22*, 605–615.

Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review*, *92*(4), 1218–1221.

Cheverst, K., Dix, A., Fitton, D., Rouncefield, M., & Graham, C. (2007). Exploring awareness related messaging through two situated-display-based systems. *Human-Computer Interaction*, *22*(1), 173–220.

Chuah, S. H., Hoffman, R., Martin, J., & Williams, G. (2007). Do cultures clash? Evidence from cross-national ultimatum game experiments. *Journal of Economic Behavior & Organization*, *55*, 35–48.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In: L. B. Resnick, J. Levine, & S. D. Teasley (Eds.) *Perspectives on Socially Shared Cognition*, (pp. 127–149). Washington DC: American Psychological Association.

Clark, H. H., & Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259–294.

Cohn, A., Fehr, E., & Herrma, B. (2009). Social comparison in the workplace: evidence from a field experiment. Mimeo, Institute for Empirical Research in Economics, University of Zurich.
URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1778894

Collett, P. (1977). The rules of conduct. In: P. Collet (Ed.) *Social rules and social behaviour*. Oxford: Blackwell.

Commons Select Committee Report (2013). MPs publish report on Google's tax avoidance. UK Government report. Available online.
URL: http://www.parliament.uk/business/committees/committees-a-z/commons-select/public-accounts-committee/news/tax-avoidance-google/

Crabtree, A., Rodden, T., & Benford, S. (2005). Moving with the times: IT research and the boundaries of CSCW. *Computer Supported Cooperative Work*, *14*, 217–251.

Croson, R. (1996). Information in ultimatum games: an experimental study. *Journal of Economic Behavior & Organization*, *30*, 197–212.

Croson, R. (2005). The method of experimental economics. *International Negotiation*, *10*, 131–148.

Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, *51*, 143–159.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. UK: Pearson: Prentice-Hall.

Davis, J. P., Farnham, S., & Jensen, C. (2002). Decreasing online 'bad' behavior. In: *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, (pp. 718–719). New York, NY, USA: ACM.

de Jong, S., Tuyls, K., & Verbeeck, K. (2008). Artificial agents learning human fairness. In: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 2*, AAMAS '08, (pp. 863–870). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Dell, N., Vaidyanathan, V., Medhi, I., Cutrell, E., & Thies, W. (2012). "Yours is better!": Participant response bias in HCI. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (pp. 1321–1330). New York, NY, USA: ACM.

Denning, P. J., & Yaholkovsky, P. (2008). Getting to "We". *Communications of the ACM*, *51*(4), 19–24.

Deutsch, M. (1975). Equity, equality and need: what determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, *31*, 137–149.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497–509.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1993). *Human-Computer Interaction*. London: Prentice-Hall.

Dong, W., & Fu, W.-T. (2012). One piece at a time: Why video-based communication is better for negotiation and conflict resolution. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, (pp. 167–176). New York, NY, USA: ACM.

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In: *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work*, CSCW '92, (pp. 107–114). New York, NY, USA: ACM.

Druskat, V. P., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, *81*(1), 58–74.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). "Alone together?": Exploring the social dynamics of massively multiplayer online games. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 407–416). New York, NY, USA: ACM.

Dufwenberg, M., & Muren, A. (2006). Generosity, anonymity, gender. *Journal of Economic Behavior & Organization*, *61*(1), 42–49.

Duggan, G. B., & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, (pp. 39–48). New York, NY, USA: ACM.

Eason, K. (1996). Division of labour and the design of systems for computer support for cooperative work. *Journal of Information Technology*, *11*, 39–50.

Eckel, C., & Grossman, P. (1996). Altruism in anonymous dictator games. *Games & Economic Behaviour*, *16*, 181–191.

Eckel, C., & Grossman, P. (2000). Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. *Experimental Economics*, *3*(2), 107–120.

Eckel, C., & Grossman, P. (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry*, *39*, 171–188.

Ellis, C. A., Gibbs, S. J., & Rein, G. (1991). Groupware: some issues and experiences. *Communications of the ACM*, *34*(1), 39–58.

Elo, S., & Kyngas, H. (2007). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*, 583–610.

Erickson, T., Halverson, C., Kellogg, W. A., Laff, M., & Wolf, T. (2002). Social translucence: designing social infrastructures that make collective activity visible. *Communications of the ACM*, *45*(4), 40–44.

Erickson, T., & Kellogg, W. A. (2000). Social translucence: an approach to designing systems that support social processes. *ACM ToCHI*, *7*, 59–83.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*, 63–87.

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

Feldman, D. C. (1984). The development and enforcement of group norms. *Academy of Management Review*, *9*, 47–53.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Fidel, R., Pejtersen, A. M., Cleal, B., & Bruce, H. (2004). A multidimensional approach to the study of human-information interaction: A case study of collaborative information retrieval. *Journal of American Society for Information Science and Technology*, *55*(11), 235–247.

Flanagin, A., Carey, L., Fontanarosa, P. B., Philips, S. G., Pace, B. P., Lundberg, G. D., & Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *Journal of the American Medical Association*, *280*(3), 222–224.

Fleck, R., Rogers, Y., Yuill, N., Marshall, P., Carr, A., Rick, J., & Bonnett, V. (2009). Actions speak loudly with words: unpacking collaboration around the table. In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, (pp. 189–196). New York, NY, USA: ACM.

Foley, C., & Smeaton, A. (2010). Division of labour and sharing of knowledge for synchronous collaborative information retrieval. *Information Processing & Management*, *46*, 762–772.

Foley, C., & Smeaton, A. F. (2009). Synchronous collaborative information retrieval: Techniques and evaluation. In: *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ECIR '09, (pp. 42–53). Berlin, Heidelberg: Springer-Verlag.

Foo, M.-D., Sin, H.-P., & Yiong, L.-P. (2006). Effects of team inputs and intrateam processes on perceptions of team viability and member satisfaction in nascent ventures. *Strategic Management Journal*, *27*, 389–399.

Forsythe, R. L., Horowitz, J., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behaviour*, *6*(3), 347–369.

Franzen, A., & Pointner, S. (2013). The external validity of giving in the dictator game: A field experiment using the misdirected letter technique. *Experimental Economics*, *16*, 155–169.

Freidson, E. (1976). The division of labor as social interaction. *Social Problems*, *23*(3), 304–313.

Frey, B., & Bohnet, I. (1995). Institutions affect fairness: Experimental investigations. *Journal of Institutional and Theoretical Economics*, *151*(2), 286–303.

Frey, B., & Bohnet, I. (1997). Identification in democratic society. *Journal of Socio-Economics*, *26*, 25–38.

Friedman, B., Kahn, P. H., , & Borning, A. (2006). Value sensitive design and information systems. In: P. Zhang, & D. Galletta (Eds.) *Human-Computer Interaction and Management Information Systems: Foundations*, (pp. 348–372). Armonk, NY: M.E. Sharpe.

Friedman, B., & Kahn, P. H. (2003). Human values, ethics, and design. In: J. A. Jacko, & A. Sears (Eds.) *The Human-Computer Interaction Handbook*, (pp. 1177–1201). Mahwah: Erlbaum.

Frost, J. (2013). What are the effects of multicollinearity and when can i ignore them? Available online. URL: http://blog.minitab.com/blog/adventures-in-statistics

Galegher, J., & Kraut, R. (1994). Computer-mediated communication for intellectual teamwork: An experiment in group writing. *Information Systems Research*, *5*(2), 110–138.

Galegher, J., & Kraut, R. E. (1990). Computer-mediated communication for intellectual teamwork: a field experiment in group writing. In: *Proceedings of the 1990 ACM conference on Computer-Supported Cooperative Work*, CSCW '90, (pp. 65–78). New York, NY, USA: ACM.

Gaver, W. W. (1991). Sound support for collaboration. In: *Proceedings of the second European Conference on Computer-Supported Cooperative Work*, ECSCW'91, (pp. 293–308). Norwell, MA, USA: Kluwer Academic Publishers.

Gaver, W. W. (1992). The affordances of media spaces for collaboration. In: *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, (pp. 17–24). New York, NY, USA: ACM.

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology*, *76*(1), 129–142.

Gintis, H. (2011). The future of behavioral game theory. *Mind & Society*, *10*(2), 97–102.

Gneezy, U., Haruvy, E., & Roth, A. E. (2003). Bargaining under a deadline: Evidence from the reverse ultimatum game. *Games and Economic Behaviour*, *45*(2), 347–368.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. London: Penguin.

Golovchinsky, G., Adcock, J., Pickens, J., Pernilla, Q., & Back, M. (2008a). Cerchiamo: a collaborative exploratory search tool. In: *Proceedings of the ACM conference on computer-supported cooperative work*, Demo Session. San Diego, CA, USA: ACM.

Golovchinsky, G., Diriye, A., & Dunnigan, T. (2012). The future is in the past: Designing for exploratory search. In: *Proceedings of the 4th Information Interaction in Context Symposium*, IIIX '12, (pp. 52–61). New York, NY, USA: ACM.

Golovchinsky, G., Diriye, A., & Pickens, J. (2011). Designing for collaboration in information seeking. In: *Proc. HCIR*.

Golovchinsky, G., & Pickens, J. (2009). A model of collaborative search. Available online. URL: http://www.slideshare.net/geneg/a-model-of-collaborative-search

Golovchinsky, G., Pickens, J., & Back, M. (2008b). A taxonomy of collaboration in online information seeking. In: *Proc. of JCDL Workshop on Collaborative Exploratory Search. Pittsburgh, PA.*.

González-Ibáñez, R., Haseki, M., & Shah, C. (2012). Time and space in collaborative information seeking. In: *Proc. ASIST*, (pp. 1–10).

González-Ibáñez, R., Haseki, M., & Shah, C. (2013). An analysis of communication and performance in collaborative information seeking. *Information Processing & Management*, *49*, 1165–1179.

González-Ibáñez, R., & Shah, C. (2011). Coagmento: A system for supporting collaborative information seeking. In: *Proc. ASIS&T*, (pp. 1–4).

Greenberg, S., & Rounding, M. (2001). The notification collage: posting information to public and personal displays. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, (pp. 514–521). New York, NY, USA: ACM.

Griffin, B. W. (2009). Cronbach's alpha, a measure of internal consistency. Available Online.
URL: http://www.bwgriffin.com/gsu/courses/edur9131/content/cronbach/cronbachs_alpha_spss.htm

Grimm, V., & Mengel, F. (2012). An experiment on learning in a multiple games environment. *Journal of Economic Theory*, *147*(6), 2220–2259.

Grinter, R. (1996). Supporting articulation work using software configuration management systems. *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, *5*, 447–465.

Grinter, R. E., Herbsleb, J. D., & Perry, D. E. (1999). The geography of coordination: dealing with distance in R&D work. In: *Proceedings of the international ACM SIGGROUP conference on supporting group work*, (pp. 306–315). New York, NY, USA: ACM.

Grudin, J. (1994a). Computer-supported cooperative work: History and focus. *IEEE Computer*, *27*(5), 19–26.

Grudin, J. (1994b). Groupware and social dynamics: eight challenges for developers. *Communications of the ACM*, *37*(1), 92–105.

Grudin, J., & Poltrock, S. (2012). Taxonomy and theory in computer supported cooperative work. In: S. W. Kozlowski (Ed.) *Handbook of Organizational Psychology*, (pp. 1323–1348). Oxford University Press.

Güth, W., Schmidt, K., & Sutter, M. (2007). Bargaining outside the lab: A newspaper experiment of a three-person ultimatum game. *The Economic Journal*, *117*(518), 449–469.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*, 367–388.

Güth, W., & Tietz, R. (1986). Auctioning ultimatum bargaining positions. In: *Issues in West German Decision Research*, (pp. 1–15). Frankfurt: Lang.

Guth, W., & Tietz, R. (1988). Ultimatum bargaining for a shrinking cake: An experimental analysis. In: R. Tietz, W. Albers, & R. Selten (Eds.) *Lecture notes in economics and mathematical systems: Bounded rational behaviour in experimental games and markets*, (pp. 111–128). Berlin: Springer.

Güth, W., & Van Damme, E. (1998). Information, strategic behaviour and fairness in ultimatum bargaining. *Journal of Mathematical Psychology*, *42*, 227–247.

Gutwin, C., & Greenberg, S. (1997). Effects of awareness support on groupware usability. Report 1997-605-07, Department of Computer Science, University of Calgary, Alberta, Canada.

Gutwin, C., & Greenberg, S. (2000). The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces. In: *Proceedings of the 9th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, WET ICE'00, (pp. 98–103). Maryland, USA: IEEE Press.

Gutwin, C., & Greenberg, S. (2001). The importance of awareness for team cognition in distributed collaboration. Report 2001-696-19, Department of Computer Science, University of Calgary, Alberta, Canada.

Gutwin, C., & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work*, *11*(3), 411–446.

Gutwin, C., Greenberg, S., & Roseman, M. (1996a). Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation. In: M. A. Sasse, R. J. Cunningham, & R. L. Winder (Eds.) *People and Computers XI*, (pp. 281–298). Springer London.

Gutwin, C., Greenberg, S., & Roseman, M. (1996b). Workspace awareness support with radar views. In: *Conference Companion on Human Factors in Computing Systems*, CHI '96, (pp. 210–211). New York, NY, USA: ACM.

Gutwin, C., Penner, R., & Schneider, K. (2004). Group awareness in distributed software development. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, (pp. 72–81). New York, NY, USA: ACM.

Hackman, J. R., Jones, L. E., & McGrath, J. E. (1967). A set of dimensions for describing the general properties of group-generated written passages. *Psychological Bulletin*, *67*(6), 379.

Hackman, R. (1988). Flight crew questionnaire. Unpublished manuscript.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*, 245–256.

Halvey, M., Vallet, D., Hannah, D., Feng, Y., & Jose, J. M. (2010). An asynchronous collaborative search system for online video search. *Information Processing & Management*, *46*(6), 733–748.

Hargittai, E. (2006). Hurdles to information seeking: Explaining spelling and typographical mistakes in users' online search behavior. *Journal of the Association of Information Systems*, *7*(1), 52–67.

Harper, F. M., Li, S. X., Chen, Y., & Konstan, J. A. (2007). Social comparisons to motivate contributions to an online community. In: *Proceedings of the 2nd international conference on Persuasive technology*, (pp. 148–159). Berlin, Heidelberg: Springer-Verlag.

Harper, R. (1992). Looking at ourselves: an examination of the social organisation of two research laboratories. In: *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, (pp. 330–337). New York, NY, USA: ACM.

Harper, R. (1996). Why people do and don't wear active badges: a case study. *Journal of Computer-Supported Cooperative Work*, *4*, 297–318.

Harper, R. R., Hughes, J. A., & Shapiro, D. Z. (1991). Harmonious working and CSCW: computer technology and air traffic control. In: J. M. Bowers, & S. D. Benford (Eds.) *Studies in Computer Supported Cooperative Work: Theory, Practice and Design*, (pp. 225–234). Amsterdam: North Holland: Elsevier Science Publishers.

Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., & Rogers, Y. (2009). Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions? In: *Proceedings of the 9th international conference on Computer Supported Collaborative Learning*, (pp. 335–344). International Society of the Learning Sciences.

Harris, R. J., & Joyce, M. A. (1980). What's fair? it depends on how you phrase the question. *Journal of Personality and Social Psychology*, *38*, 165–170.

Harrison, D. A., Mohammed, S., McGrath, J. E., Florey, A. T., & Vanderstoep, S. W. (2003). Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality. *Personnel Psychology*, *56*(3), 633–669.

Harrison, G. W., & List, J. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Heath, C., Jirotka, M., Luff, P., & Hindmarsh, J. (1995). The individual and the collaborative: The interactional organisation of trading in a city dealing room. *Computer Supported Cooperative Work*, *3*(1), 147–165.

Heath, C., & Luff, P. (1992). Collaboration and control: Crisis management and multimedia technology in london underground line control rooms. *Computer Supported Cooperative Work*, *1*(1-2), 147–165.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4*, 5–28.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford, UK: Oxford University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of *Homo Economicus*: Behavioral experiments in 15 small-scale societies. In: *The American Economic Review, Vol. 91, No. 2, Papers and Proceedings of the Hundred Thirteen Annual Meeting of the American Economic Association*, (pp. 73–78). USA: American Economic Association.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(29), 29.

Herbsleb, J., & Grinter, R. (1999). Architectures, coordination, and distance: Conway's law and beyond. *Software, IEEE*, *16*(5), 63–70.

Hertel, G., Aarts, H., & Zeelenberg, M. (2002). What do you think is 'fair'? Effects of ingroup norms and outcome control on fairness judgments. *European Journal of Social Psychology*, *32*, 327–341.

Hertel, G., Deter, C., & Konradt, U. (2003). Motivation gains in computer-supported groups. *Journal of Applied Social Psychology*, *33*(10), 2080–2105.

Hobbs, A. (2010). What is fairness?
URL: http://news.bbc.co.uk/today/hi/today/newsid_9079000/9079254.stm

Hochheiser, H., & Lazar, J. (2007). HCI and societal issues: A framework for engagement. *International Journal of Human-Computer Interaction*, *23*(3), 339–374.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. L. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behaviour*, *7*(1), 346–380.

Hoffman, E., McCabe, K., & Smith, V. L. (1996a). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, *25*, 289–301.

Hoffman, E., McCabe, K., & Smith, V. L. (1996b). Social distance and other-regarding behaviour in dictator games. *American Economic Review*, *86*(3), 653–660.

Hoffman, E., McCabe, K., & Smith, V. L. (2008). Preferences and property rights in ultimatum and dictator games. In: *Handbook of Experimental Economics Results*, vol. 1, (pp. 417–422). Elsevier Science Publishers.

Holand, U., & Danielsen, T. (1991). Describing cooperation – the creation of different psychological phenomena. In: *Studies in Computer Supported Cooperative Work: Theory, Practice and Design*, (pp. 17–27). Amsterdam: North Holland: Elsevier Science Publishers.

Holt, H., & Thaulow, I. (1996). Formal and informal flexibility in the workplace. *The work-family challenge: Rethinking employment*, (pp. 79–82).

Hornecker, E., & Nicol, E. (2012). What do lab-based user studies tell us about in-the-wild behavior?: insights from a study of museum interactives. In: *Proceedings of the Designing Interactive Systems Conference*, DIS '12, (pp. 358–367).

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, *116*(4), 717–751.

Hudson, S. E., & Smith, I. (1996). Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In: *Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work*, (pp. 248–257). New York, NY, USA: ACM.

Hulin, C. (2001). Measurement: Cronbach's alpha on two-item scales. *Journal of Consumer Psychology*, *10*(1 & 2), 55–69.

Isen, A. M. (1970). Success, failure, attention, and reaction to others: The warm glow of success. *Journal of Personality and Social Psychology*, *15*(4), 294–301.

Jackson, J. M., & Harkins, S. G. (1985). Equity in effort: An explanation of the social loafing effect. *Journal of Personality and Social Psychology*, *49*(5), 1199–1206.

Jang, C.-Y. (2007). *Managing Fairness: Reward Distribution in a Self-organized Online Game Player Community*. Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag.

Jensen, C., Farnham, S. D., Drucker, S. M., & Kollock, P. (2000). The effect of communication modality on cooperation in online environments. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, (pp. 470–477). New York, NY, USA: ACM.

Johansen, R., Sibbet, D., Benson, S., Martin, A., Mittman, R., & Saffo, P. (1991). *Leading business teams*. Reading, MA: Addison-Wesley.

Judd, C. M., & McLelland, G. H. (1989). *Data analysis: A model comparison approach*. San Diego, CA: Harcourt Brace Jovanovich.

Kagel, J. H., Kim, C., & Moser, D. (1996). Fairness in ultimatum games with asymmetric information and asymmetric payoffs. *Games and Economic Behavior*, *13*, 100–110.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986a). Fairness and the assumptions of economics. *Journal of Business*, *59*(4), 285–300.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986b). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, *76*, 728–741.

Kang, R., Fu, W.-T., & Kannampallil, T. G. (2010). Exploiting knowledge-in-the-head and knowledge-in-the-social-web: Effects of domain expertise on exploratory search in individual and social search environments. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, (pp. 393–402). New York, NY, USA: ACM.

Karau, S. J., & Kelly, J. R. (1992). The effects of time scarcity and time abundance on group performance quality and interaction process. *Journal of Experimental Social Psychology*, *28*(6), 542–571.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*(4), 681–706.

Kazakos, K., Howard, S., & Vetere, F. (2013). Revisiting the relationship between reunion and technology-mediated separation in periodically transitioning families. In: *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, (pp. 1157–1168). New York, NY, USA: ACM.

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, *3*(1), 1–224.

Kelly, J. R., Futoran, G. C., & McGrath, J. E. (1990). Capacity and capability: Seven studies of entrainment of task performance rates. *Small Group Research*, *21*(3), 283–314.

Kelly, J. R., & Karau, S. J. (1999). Group decision making: The effects of initial preferences and time pressure. *Personality and Social Psychology Bulletin*, *25*(11), 1342–1354.

Kelly, J. R., & Loving, T. J. (2004). Time pressure and group performance: Exploring underlying processes in the attentional focus model. *Journal of Experimental Social Psychology*, *40*(2), 185–198.

Kelly, J. R., & McGrath, J. E. (1985). Effects of time limits and task types on task performance and interaction of four-person groups. *Journal of Personality and Social Psychology*, *49*(2), 395.

Kelly, R., & Payne, S. J. (2013). Division of labour in collaborative information seeking: Current approaches and future directions. In: *Third international workshop on collaborative information seeking*, Hosted at the 2013 conference on Computer supported cooperative work, (pp. 1–4).

Kelly, R., & Payne, S. J. (2014). Collaborative web search in context: A study of tool use in everyday tasks. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, (pp. 807–819). New York, NY, USA: ACM.

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, *45*, 819–828.

Kiesler, S., Sproull, L., & Waters, K. (1996). A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology*, *70*, 47–66.

Kim, T., Hinds, P., & Pentland, A. (2012). Awareness as an antidote to distance: making distributed groups cooperative and consistent. In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, (pp. 1237–1246). New York, NY, USA: ACM.

Kittur, A. (2010). Crowdsourcing, collaboration and creativity. *XRDS*, *17*(2), 22–26.

Kling, R. (1993). Fair information practices with computer supported cooperative work. *SIGOIS Bulletin*, *14*(1), 28–31.

Knez, M. J., & Camerer, C. F. (1995). Outside options and social comparison in three-player ultimatum game experiments. *Games and Economic Behaviour*, *10*, 65–94.

Kolfschoten, G. L., Herrmann, T., & Lukosch, S. (2013). Differentiated awareness-support in computer supported collaborative work. *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, *22*(2–3), 107–112.

Komorita, S. S., & Parks, C. D. (1995). Interpersonal Relations: Mixed-motive interaction. *Annual Review of Psychology*, *46*, 183–207.

Kraut, R. E. (2003). Applying social psychological theory to the problems of group work. In: J. M. Carroll (Ed.) *HCI Models, Theories and Frameworks: Toward A Multidisciplinary Science*, (pp. 335–356). Amsterdam: North Holland: Elsevier Science Publishers.

Kraut, R. E., Fish, R., Root, R., & Chalfonte, B. (1990). Informal communication in organizations: Form, function and technology. In: S. Oskamp, & S. Spacapan (Eds.) *Human Reactions to Technology*, (pp. 145–199). Beverly Hills, CA: Sage Publications.

Kravitz, D. A., & Gunto, S. (1992). Decisions and perceptions of recipients in ultimatum bargaining games. *Journal of Socio-Economics*, *21*(1), 65–84.

Kravitz, D. A., & Martin, B. (1986). Ringelmann rediscovered: The original article. *Journal of Personality and Social Psychology*, *50*(5), 936–941.

Kurzban, R. (2013). Are all dictator game results artifacts? available online.
URL: http://www.epjournal.net/blog/2013/05/

Labella, T. H., Dorigo, M., & Deneubourg, J.-L. (2006). Division of labor in a group of robots inspired by ants' foraging behavior. *ACM Transactions on Autonomous and Adaptive Systems*, (pp. 4–25).

Lapadat, J. C., & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry*, *5*(1), 64–86.

Larrick, R. P., & Blount, S. (1997). The claiming effect: Why players are more generous in social dilemmas than in ultimatum games. *Journal of Personality and Social Psychology*, *72*(4), 810–825.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*, 822–832.

Lazem, S., & Gračanin, D. (2010). Social traps in second life. In: *Proceedings of the 2010 Second International Conference on Games and Virtual Worlds for Serious Applications*, VS-GAMES '10, (pp. 133–140). Washington, DC, USA: IEEE Computer Society.

Lazem, S., Gračanin, D., & Harrison, S. (2012). Effects of context-sensitive delays on group dynamics in 3d virtual worlds. In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, CSCW '12, (pp. 155–158). New York, NY, USA: ACM.

Lerner, M. J. (1971). Observer's evaluation of a victim: Justice, guilt, and veridical perception. *Journal of Personality and Social Psychology*, *20*(2), 127.

Leventhal, G. (1976). Fairness in social relationships. In: J. W. Thibaut, J. T. Spence, & R. C. Carson (Eds.) *Contemporary topics in social psychology*, (pp. 211–239). Morristown, New Jersey: General Learning Press.

Leventhal, G. S. (1980). *What should be done with equity theory?*. Springer.

Levine, J. M., & Moreland, R. L. (1998). Small Groups. In: D. T. Gilbert, & S. T. Fiske (Eds.) *The Handbook of Social Psychology, Vol. 2*, (pp. 415–469). Boston, MA: Mcgraw-Hill.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*(2), 153–174.

Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin*, *30*, 1175–1185.

Lind, E. A., Kanfer, R., & Earley, P. C. (1990). Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments. *Journal of Personality and Social Psychology*, *59*, 952–959.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, *115*(3), 482–493.

List, J. A., & Cherry, T. L. (2000). Learning to accept in ultimatum games: Evidence from an experimental design that generates low offers. *Experimental Economics*, *3*, 11–31.

List, J. A., & Cherry, T. L. (2008). Examining the role of fairness in high stakes allocation decisions. *Journal of Economic Behavior & Organization*, *65*, 1–8.

Malone, T. W., & Crowston, K. (1990). What is coordination theory and how can it help design cooperative work systems? In: *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, (pp. 357–370). New York, NY, USA: ACM.

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, *26*(1), 87–119.

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, *36*, 158–160.

Mantei, M. M., Baecker, R. M., Sellen, A. J., Buxton, W. A. S., Milligan, T., & Wellman, B. (1991). Experiences in the use of a media space. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 203–208). New York, NY, USA: ACM.

Mark, G., Haake, J. M., & Streitz, N. A. (1996). Hypermedia structures and the division of labor in meeting room collaboration. In: *Proc. CSCW '96*, (pp. 170–179).

Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., & Davies, M. (2011). Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, (pp. 3033–3042).

McDonald, J. H. (2008). *Handbook of Biological Statistics*. Baltimore, USA: Sparky House Publishing.

McGrath, J. E. (1990). Time matters in groups. In: J. Galegher, R. E. Kraut, & C. Egido (Eds.) *Intellectual Teamwork: The Social and Technological Bases of Cooperative Work*, (pp. 23–61). Hillsdale, NJ: Lawrence Erlbaum.

Mengel, F. (2012). Learning across games. *Games & Economic Behaviour*, *74*(2), 610–619.

Merritt, T., & McGee, K. (2012). Protecting artificial team-mates: more seems like less. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, (pp. 2793–2802). New York, NY, USA: ACM.

Middup, C. P. (2008). *Supporting Collaboration in Problem Solving Groups*. Ph.D. thesis, University of Bath.

Mikula, G. (1998). Division of labour and perceived justice: a growing field of research. *Social Justice Research*, *11*(3), 215–241.

Miller, D., & Vidmar, N. (1981). The social psychology of punishment reactions. In: M. J. Lerner, & S. Lerner (Eds.) *The justice motive in social behavior*, (pp. 145–172). New York, NY: Plenum.

Mitzkewitz, M., & Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory*, *22*, 171–198.

Monk, A., & Watts, L. (2000). Peripheral participation in video-mediated communication. *International Journal of Human-Computer Studies*, *52*(5), 933–958.

Moore, R. J., Ducheneaut, N., & Nickell, E. (2007a). Doing virtually nothing: Awareness and accountability in massively multiplayer online worlds. *Computer Supported Cooperative Work*, *16*, 265–305.

Moore, R. J., Gathman, E. C. H., Ducheneaut, N., & Nickell, E. (2007b). Coordinating joint activity in avatar-mediated interaction. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, (pp. 21–30). New York, NY, USA: ACM.

Morris, M. R. (2008). A survey of collaborative web search practices. In: *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, (pp. 1657–1660). New York, NY, USA: ACM.

Morris, M. R. (2013). Collaborative search revisited. In: *Proc.*, CSCW '13, (pp. 1181–1192).

Morris, M. R., & Horvitz, E. (2007). Searchtogether: an interface for collaborative web search. In: *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology*, (pp. 3–12). New York, NY, USA: ACM.

Morris, M. R., Lombardo, J., & Wigdor, D. (2010). Wesearch: Supporting collaborative search and sense-making on a tabletop display. In: *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, (pp. 401–410). New York, NY, USA: ACM.

Murnighan, J. K., & Saxon, M. S. (1998). Ultimatum bargaining by children and adults. *Journal of Economic Psychology*, *19*, 415–445.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*(3), 472–489.

Nardi, B., & Harris, J. (2006). Strangers and friends: collaborative play in world of warcraft. In: *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, (pp. 149–158). New York, NY, USA: ACM.

Nardi, B. A., & Miller, J. R. (1990). An ethnographic study of distributed problem solving in spreadsheet development. In: *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, (pp. 197–208). New York, NY, USA: ACM.

Neale, D. C., Carroll, J. M., & Rosson, M. B. (2004). Evaluating computer-supported cooperative work: models and frameworks. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, (pp. 112–121).

Nelson, W. R. (2002). Equity or intention: It is the thought that counts. *Journal of Economic Behavior & Organization*, *48*, 423–430.

Nguyen, D. T., & Canny, J. (2007). Multiview: improving trust in group video conferencing through spatial faithfulness. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1465–1474). New York, NY, USA: ACM.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.

Ochs, J., & Roth, A. E. (1989). An experimental study of sequential bargaining. *American Economic Review*, *79*, 355–384.

Ockenfels, A., & Werner, P. (2012). Hiding behind a small cake in a newspaper dictator game. *Journal of Economic Behavior & Organization*, *82*, 82–85.

Okada, K.-I. (2007). Collaboration support in the information sharing space. *IPSJ Magazine*, *48*(2), 123–125.

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction*, *15*, 139–178.

Olson, G. M., Olson, J. S., Carter, M. R., & Storrøsten, M. (1992a). Small group design meetings: An analysis of collaboration. *Human-Computer Interaction*, *7*(4), 347–374.

Olson, J. S., Olson, G. M., Storrøsten, M., & Carter, M. (1992b). How a group-editor changes the character of a design meeting as well as its outcome. In: *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, (pp. 91–98). New York, NY, USA: ACM.

O'Neill, J., & Martin, D. (2013). Relationship-based business process crowdsourcing? In: *In proceedings of the 14th IFIP Conference on Human-Computer Interaction*, INTERACT '13, (pp. 429–446). Berlin, Heidelberg: Springer.

Oosterbeek, H., Sloof, R., & van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, *7*, 171–188.

Opp, K. D. (1982). The evolutionary emergence of social norms. *British Journal of Social Psychology*, *21*, 139–149.

Orlikowski, W. J. (1992). Learning from notes: organizational issues in groupware implementation. In: *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, CSCW '92, (pp. 362–369). New York, NY, USA: ACM.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*(6).

Oulasvirta, A., Hukkinen, J., & Schwartz, B. (2009). When more is less: The paradox of choice in search engine use. In: *Proceedings of SIGIR*, (pp. 516–523). New York, NY, USA: ACM.

Oxoby, R. J., & Spraggon, J. (2006). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, *65*(3), 703–713.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1996). My partner is a real dog: cooperation with social agents. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, CSCW '96, (pp. 399–408).

Park, S.-T., & Pennock, D. M. (2007). Applying collaborative filtering techniques to movie search for better ranking and browsing. In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, (pp. 550–559).

Paul, S. A., & Morris, M. R. (2009). Cosense: enhancing sensemaking for collaborative web search. In: *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, (pp. 1771–1780). New York, NY, USA: ACM.

Paul, S. A., & Morris, M. R. (2011). Sensemaking in collaborative web search. *Human-Computer Interaction*, *26*(1-2), 72–122.

Perla, P. P., Markowitz, M., Nofi, A., Weuve, C., Loughran, J., & Stahl, M. (2000). Gaming and shared situation awareness. Naval Analyses Report CRM D0002722.A2.

Pinelle, D., Gutwin, C., & Greenberg, S. (2003). Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *ACM Transactions on Computer-Human Interaction*, *10*(4), 281–311.

Plonka, L., Segal, J., Sharp, H., & Linden, J. v. d. (2012). Investigating equity of participation in pair programming. In: *Proceedings of Agile India 2012*, (pp. 20–29). Washington, DC, USA.

Plowman, L., Rogers, Y., & Ramage, M. (1995). What are workplace studies for? In: *Proc.*, ECSCW'95, (pp. 309–324). Norwell, MA, USA: Kluwer Academic Publishers.

Poltrock, S., Grudin, J., Dumais, S., Fidel, R., Bruce, H., & Pejtersen, A. M. (2003). Information seeking and sharing in design teams. In: *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, GROUP '03, (pp. 239–247). New York, NY, USA: ACM.

Poundstone, W. (1992). *The Prisoner's Dilemma*. New York, NY: Doubleday.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*(5), 1281–1302.

Raghubir, P., & Srivastava, J. (2002). Effect of face value on product valuation in foreign currencies. *Journal of Consumer Research*, *29*, 335–347.

Ramage, M. (1994). *Engineering a smooth flow? A study of workflow software and its connections with business process reengineering*. Master's thesis, University of Sussex, United Kingdom.

Rapoport, A., Sundali, J. A., & Potter, R. E. (1996). Ultimatums in two-person bargaining with one-sided uncertainty. *International Journal of Game Theory*, *25*, 475–494.

Rawls, J. (1999). *A Theory of Justice*. Cambridge, MA: Belknap.

Reed, D., McGee, D., Katsuhiko, Y., & Hankin, J. (1985). Diet, blood pressure, and multicollinearity. *Hypertension*, *7*(3), 405–410.

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Stanford, CA: CSLI Publications.

Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2003). The researcher's dilemma: Evaluating trust in computer-mediated communication. *International Journal of Human-Computer Studies*, *58*(6), 759–781.

Riessman, C. K. (1993). *Narrative Analysis*. Newbury Park, CA: Sage.

Roberts, N. C. (1991). Stakeholder collaboration and innovation: A study of public policy initiation at the state level. *Journal of Applied Behavioral Science*, *27*(2), 209–227.

Rocco, E. (1998). Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, (pp. 496–502). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co.

Rodden, T. (1991). A survey of CSCW systems. *Interacting with Computers*, *3*(3), 319–353.

Rogers, Y. (1993). Coordinating computer-mediated work. *Computer Supported Cooperative Work*, *1*(4), 295–315.

Rogers, Y. (2011). Interaction design gone wild: striving for wild theory. *interactions*, *18*(4), 58–62.

Rogers, Y., Lim, Y.-K., Hazlewood, W. R., & Marshall, P. (2008). Equal opportunities: Do shareable interfaces promote more group participation than single users displays? *Human-Computer Interaction*, *24*(1-2), 79–116.

Rosas, J., & Camarinha-Matos, L. M. (2009). An approach to assess collaboration readiness. *Journal of Production Research*, *47*(17), 4711–4735.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In: C. O'Malley (Ed.) *Computer-Supported Collaborative Learning*, (pp. 69–97). Springer Verlag.

Rosenthal, R. W., & Rosnow, R. L. (1969). *Artifact in Behavioural Research*. New York: Academic Press.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behaviour in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review*, *81*, 245–261.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958–1992. *Rationality and Society*, *7*(1), 58–92.

Sandelowski, M. (1995). Qualitative analysis: what it is and how to begin? *Research in Nursing & Health*, *18*, 371–375.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758.

Schmidt, K. (1994). The organization of cooperative work: Beyond the 'Leviathan' conception of the organization of cooperative work. In: *Proc. CSCW '94*, (pp. 101–112).

Schmidt, K. (2000). The critical role of workplace studies in CSCW. In: C. Heath, J. Hindmarsh, & P. Luff (Eds.) *Workplace Studies: Recovering Work Practice and Informing Design*, (pp. 141–149). Cambridge University Press.

Schmidt, K. (2002). The problem with 'awareness'. *Computer Supported Cooperative Work*, *11*, 285–298.

Schmidt, K. (2010). Keep up the good work!: The concept of 'work' in cscw. In: *Proceedings of the 9th International Conference on Designing Cooperative Systems*, (pp. 265–286). Berlin, Heidelberg: Springer-Verlag.

Schmidt, K. (2011). *Cooperative Work and Coordinative Practices*. Springer-Verlag.

Schmidt, K., & Bannon, L. (1992). Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work*, *1*(1–2), 7–40.

Schweitzer, M., & Solnick, S. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes*, *79*, 199–215.

Scissors, L. E., & Gergle, D. (2013). "Back and forth, back and forth": Channel switching in romantic couple conflict. In: *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, (pp. 237–248). New York, NY, USA: ACM.

Scissors, L. E., Gill, A. J., Geraghty, K., & Gergle, D. (2009). In CMC we trust: the role of similarity. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, (pp. 527–536). New York, NY, USA: ACM.

Scissors, L. E., Gill, A. J., & Gergle, D. (2008). Linguistic mimicry and trust in text-based CMC. In: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, (pp. 277–280). New York, NY, USA: ACM.

Segal, L. D. (1994). Actions speak louder than words: How pilots use nonverbal information for crew communications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *38*(1), 21–40.
URL: http://pro.sagepub.com/content/38/1/21.abstract

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, *4*, 25–55.

Shafir, E., Diamond, P., & Tversky, A. (1997). The money illusion. *Quarterly Journal of Economics*, *112*, 342–374.

Shah, C. (2010). Collaborative information seeking: A literature review. *Advances in Librarianship*, *32*, 1–31.

Shah, C., & González-Ibáñez, R. (2010). Spatial context in collaborative information seeking. *Journal of Information Science*, *38*(4), 33–349.

Shah, C., & González-Ibáñez, R. (2011). Evaluating the synergic effect of collaboration in information seeking. In: *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, (pp. 913–922). New York, NY, USA: ACM.

Shah, C., & Marchionini, G. (2010). Awareness in collaborative information seeking. *Journal of American Society for Information Science and Technology*, *61*(10), 1970–86.

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*(9), 803–809.

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In: R. Srull, & T. Wyer (Eds.) *Handbook of social cognition (Vol.1)*, (pp. 111–128). Hillsdale, NJ: Erlbaum.

Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures*. Chapman.

Slonim, R., & Roth, A. (1998). Learning in high stakes ultimatum games: An experiment in the slovak republic. *Econometrica*, *66*(3), 569–596.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, *66*(2), 274–279.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, *72*(5), 923–955.

Solnick, S. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, *39*, 189–200.

Stokes, J. P. (1983). Components of group cohesion: Intermember attraction, instrumental value, and risk taking. *Small Group Behavior*, *14*, 163–173.

Stoop, J. (2012). From the lab to the field: envelopes, dictators and manners. Munich Personal RePEc Archive, MPRA Paper.
URL: http://mpra.ub.uni-muenchen.de/37048/

Straub, P. G., & Murnighan, J. K. (1995). An experimental investigation of ultimatum games: Common knowledge, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior & Organization*, *27*, 345–364.

Straus, S. G. (1999). Testing a typology of tasks: An empirical validation of McGraths (1984) group task circumplex. *Small Group Research*, *30*(2), 166–187.

Strauss, A. (1985). Work and the division of labor. *The Sociological Quarterly*, *26*(1), 1–19.

Strauss, A., & Corbin, J. M. (1998). *Basics of Qualitative Research*. Sage.

Strauss, A., Fagerhuagh, S., Suczek, B., & Claude, W. (1995). *The Social Organisation of Medical Work*. Chicago: Chicago University Press.

Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge: Cambridge University Press.

Symon, G., Long, K., & Ellis, J. (1996). The coordination of work activities: cooperation and conflict in a hospital context. *Journal of Computer Supported Cooperative Work*, *5*(1), 1–31.

Tabachnik, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. New York, USA: Allyn & Bacon.

Takahashi, T. (2007). Economic decision-making in the ultimatum game by smokers. *Neuro Endocrinology Letters*, *28*(5), 659–661.

Talmon, Y. (1972). *Family and Community in the Kibbutz*. Boston, MA: Harvard University Press.

Tang, J. C. (1991). Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, *34*(2), 143–160.

Tao, Y., & Tombros, A. (2013). An exploratory study of sensemaking in collaborative information seeking. In: *Advances in Information Retrieval*, (pp. 26–37). Springer.

Tatarkiewicz, W. (1976). *Analysis of Happiness*. The Hague; The Netherlands: Martinus Nijhoff.

Thaler, R., & Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Penguin.

Thompson, L. L. (1996). Negotiation: Overview of theory and research. In: L. L. Thompson (Ed.) *Negotiation Theory and Research*, (pp. 1–6). New York, Hove: Psychology Press.

Tomasello, M. (2009). *Why we cooperate*. Cambridge, MA: MIT Press.

Torkel, F. E. (2007). Building a better damage meter. Available online.
URL: http://beastwith.in/2007/07/12/building-a-better-damage-meter/

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *221*(4481), 453–458.

Tversky, A., Shmuel, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*, 371–384.

van den Bos, K., Lind, E. A., Vermunt, R., & Wilke, H. A. M. (1997). How do I judge my outcome when I do not know the outcomes of others? The psychology of the fair process effect. *Journal of Personality and Social Psychology*, *72*(5), 1034–1046.

Voorhees, E. M. (2006). Overview of the TREC 2005 robust retrieval track. In: *Proceedings of TREC-14*.

Wallace, J. R., Scott, S. D., & MacGregor, C. G. (2013). Collaborative sensemaking on a digital tabletop and personal tablets: prioritization, comparisons, and tableaux. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 3345–3354). New York, NY, USA: ACM.

Wertenbroch, K., Soman, D., & Chattopadhyay, A. (2007). On the perceived value of money: The reference dependence of currency numerosity effects. *Journal of Consumer Research*, *34*(1), 1–10.

Wierzbicka, A. (2006). *English: Meaning and Culture*. London, UK: Oxford University Press.

Wilkinson, N. (2008). *An Introduction to Behavioral Economics*. Palgrave Macmillan.

Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., & Nickell, E. (2006). From tree house to barracks: The social life of guilds in world of warcraft. *Games and Culture*, *1*, 338–361.

Williams, K. D., Harkins, S., & Latane, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology*, *40*, 303–311.

Wilson, B. J., Abbate, P. D., Bogart, A. B., Buchanan, J., Gamboa, M. A., McMahon, M. K., & Simpson, M. (2010). Smile, dictator, you're on camera. Working paper.
URL: http://www1.chapman.edu/~bjwilson/papers/StreetDictator.pdf

Wilson, P. (1991). *Computer Supported Cooperative Work: An Introduction*. Kluwer Academic Publishers.

Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behaviour*, (pp. 288–293).

Wong, N., & Gutwin, C. (2010). Where are you pointing?: The accuracy of deictic pointing in CVEs. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1029–1038). New York, NY, USA: ACM.

Yee, N. (2006a). The psychology of massively multi-user online role-playing games: Motivations, emotional investment, relationships and problematic usage. In: R. Schroeder, & A. S. Axelsson (Eds.) *Avatars at Work and Play*, (pp. 187–207). Springer.

Yee, N. (2006b). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and Virtual Environments*, *15*(3), 309–329.

Yee, N. (2006c). The labor of fun: How video games blur the boundaries of work and play. *Games and Culture*, *1*, 68–71.

Yee, N., & Bailenson, J. (2007). The Proteus Effect: Behavioral Modification via Transformations of Digital Self-Representation. *Human Communication Research*, *33*, 271–290.

Zaccaro, S. J., & McCoy, M. C. (1988). The effects of task and interpersonal cohesiveness on performance of a disjunctive group task. *Journal of Applied Social Psychology*, *18*, 837–851.

Zheng, J., Bos, N., Olson, J. S., & Olson, G. M. (2001). Trust without touch: jump-start trust with social chat. In: *Extended Abstracts on Human Factors in Computing Systems*, (pp. 293–294). New York, NY, USA: ACM.

Zheng, J., Veinott, E., Bos, N., Olson, J. S., & Olson, G. M. (2002). Trust without touch: jumpstarting long-distance trust with initial social activities. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 141–146). New York, NY, USA: ACM.

Zizzo, D. J. (2010). Experimental demand effects in economic experiments. *Experimental Economics*, *13*, 75–98.

Appendices

Appendix A

Ethics Checklist

Contents: Department of Computer Science 13-point ethics checklist, and our responses.

A.1 Completed Ethics Checklist: Overview

This checklist was reviewed prior to each experiment. The overview we give here is retrospective to show how particular issues were handled. An official copy of this checklist can be found at http://www.cs.bath.ac.uk/Leon/files/EthicsChecklist.pdf.

1. *Have you provided a briefing script for volunteers?*

   All of our studies had briefing scripts that describe the purpose of the study. All participants are aware that they can terminate their participation without consequence.

2. *Will the participants be using any non-standard hardware?*

   No. Participants in study 6 used non-standard software applications with which they were unfamiliar. However, these tools pose no immediate physical or mental risks to participants. This is true of all other software used in this thesis.

3. *Is there any intentional deception of participants?*

   No. Our DLUG studies did not use deception but the precise purpose of the work was withheld until after participants had completed the experiments. This was necessary to prevent knowledge of our focus impacting behavioural decisions—such knowledge would invalidate the findings of the studies. The study was presented to participants as a study of division of labour in collaborative information seeking, a statement that is true but obfuscated our focus on the DLUG outcome. All participants were given full debriefing about the purpose of the study *after* they had completed the experimental task.

4. *How will participants voluntarily give consent?*

   All experimental studies used paper-based consent forms to obtain consent. Our survey study (Study 1) was conducted online, and participants gave consent on the first page of the survey. In Studies 6 and 7, consent was obtained verbally by reading the consent form aloud in cases where interviews were conducted over the telephone.

5. *Will participants be exposed to risks greater than those encountered in their ordinary working life?*

   No.

6. *Are you offering any incentive to the participants?*

   Yes. Such incentives were our primary method of recruiting volunteers, as opposed to coercion or forced participation. No participants were forced to participate in our studies. The payments we offer do not encourage participants to risk physical or mental harm beyond what they would encounter in everyday life.

7. *Are any of your participants under the age of 16?*

   No.

8. *Do any of your participants have an impairment that will limit their understanding or communication?*

   No.

9. *Are you in a position of authority over any participants?*

   In some cases, yes. This was true of Study 1, where we recruited undergraduate students that the author had previously tutored. However, we were careful to ensure that students did not feel forced to participate, as evidenced by the fact that not all students responded to our survey. Additionally, we made sure that students were aware that responses would not affect their marks. We also offered cash prizes as a thank you to the students for giving up their time to participate in our survey. Students were not requested to provide identifying information, which we hoped would put them at ease for the purposes of our survey.

10. *Will the participants be informed that they can withdraw at any time?*

    Yes. Our consent forms make this explicitly clear.

11. *Will the participants be informed of your contact details?*

    Yes. These are available at the bottom of all consent forms and instructions.

12. *Will participants be debriefed?*

    Yes. All participants received verbal (Studies 6 and 7), digital (Study 1) or paper-based debreifings (Studies 2, 3, 4, 5).

13. *Will the data collected from participants be stored in anonymous form?*

    Yes. No participants are identifiable in digital materials representing results. Additionally, paper-based materials are kept in a locked drawer inaccessible to anyone other than the author. Consent forms with identifying information are stored separately from data.

Appendix B

Study 1 Materials

Contents: Survey transcript and briefing materials.

B.1 Survey Transcript

This survey was hosted using Qualtrics survey software.[39]

Page 1—Briefing Script

Thank you for your interest in this study. This is a survey study where you will be asked to reflect on your experiences during the Integrated Project unit. The purpose of the study is to explore your satisfaction with your Integrated Project group, focusing in particular on the way in which your group organised and completed its work. The survey has 5 pages and should take roughly 10 minutes to complete.

The information below is about confidentiality and how the data we collect will be analysed. Please take a moment to read through the information before progressing to the next page.

—

During the survey you will be asked to express opinions about your coursework group. All of your responses are entirely confidential. Your group members will NOT be informed about your individual responses, and you will not be informed of theirs.

Your responses to this questionnaire will NOT impact your mark in any way. If, for example, you admit here that you did not contribute fully to the group or that someone else not participate, this information will not be used to adjust your individual or group mark in any way whatsoever. We would like your honest opinion, so please feel free to be frank, open, and honest.

There are 10 prizes of £10 on offer for completing the survey. If you wish to be included in the prize draw, you will be asked to provide your Bath email ID at the end of the survey. This is for winner identification purposes and all IDs will be removed from the dataset prior to analysis. Prize winners will be drawn at random and notified privately by email. Entry into the draw is contingent upon completion of the questionnaire.

—

If you are ready to begin, please click Next.

[39] http://www.qualtrics.com

Page 2—Demographics

As mentioned previously, we are interested in exploring your overall satisfaction with your Integrated Project group. During the survey you will be asked to express your agreement with a series of statements about your group. The scales are labelled from 'Strongly Disagree' to 'Strongly Agree'. Consider the midpoint between these as a neutral response.

We would like to stress that although we ask for your IP group number, we are primarily interested in group-level phenomena and you will not be identified as individuals during analysis or presentation of data.

Hereafter, there are only 3 pages of questions. Some questions require an answer. Optional questions have '(Optional)' written alongside them.

When you are ready, please fill in the demographic information below and proceed to the next page.

Age:

Gender (M/F):

IP Group Number:

Page 3—About your group

We would like to begin by asking you to think about your group and express your agreement with a number of statements.

Thinking about your experiences with your Integrated Project group, to what extent do you agree with the following statements?

*(Statements from the following measures were then presented in two separate blocks of 11 statements each, with order of statements randomised for each participant. Participants gave their response on a seven-point scaled, labelled at the leftmost end with 'Strongly Disagree' and 'Strongly Agree' at the rightmost. Statements with *** were reverse scored.)*

Open Communication, 4 items (Stokes, 1983)

- *My group avoids saying anything that might upset someone.****
- *My group is very straightforward with me.*
- *There are certain topics about our work that my group avoids talking about.****
- *Most people in my group are careful not to reveal too much of themselves to the group.****

Task Motivation, 4 of 6 items (Zaccaro & McCoy, 1988)

- *Performing well is a top priority for my team.*
- *My group members expect high effort and commitment from me.*

- *Only a high level of performance is acceptable to our group.*
- *Our group is highly task-oriented.*

Group Viability, 3 items (Hackman, 1988)

- *As a team, this work group shows signs of falling apart.****
- *Members of my team care a lot about it, and work together to make it one of the best.*
- *Working with members of my team is an energizing and uplifting experience.*

Group Cohesion, 5 items (Stokes, 1983)

- *If I were to participate in another group like this one, I would want it to include people who are very similar to the ones in this group.*
- *Most of the people in the group are not the kind of people I would enjoy spending time with outside the group sessions.****
- *There are not many people I like as individuals in my group.****
- *Even if we stopped meeting as a group, I would still want to see the people in this group as often as I could.*
- *I wish I had more time for socializing with other group members.*

Satisfaction With Group – (Hackman, 1988), 3 items

- *Generally speaking, I am satisfied with my team.*
- *I frequently wish I could quit the team.****
- *I am generally satisfied with the work I do on the team.*

Social Loafing – (Druskat & Wolff, 1999), 3 items

- *We have some team members that don't put much effort into their work.*
- *Every member of our team does his/her share of the work.****
- *There are some individuals on our team who dont do much work.*

If you wish to clarify, explain, or say anything else about any of your ratings on this page, please use the space below. (Optional)

Free text response box here.

Page 4—The division of work in your group

The following statements are about the way in which your group went about organising its work, and whether or not you believe the division of work was fair.

Thinking about your Integrated Project group, to what extent do you agree with each of the following statements?

*(Statements from the following measures were then presented, with order of statements randomised for each participant. Participants gave their response on a seven-point scaled, labelled at the leftmost end with 'Strongly Disagree' and 'Strongly Agree' at the rightmost. Statements with *** were reverse scored.)*

Fairness in Division of Labour, 3 items

- *My team made an effort to ensure that work was distributed fairly among members.*
- *In my opinion, the division of work among my group members was fair.*
- *When dividing up tasks, some members of my group were given an unfair amount of work.****

Fairness in Completion of Work, 4 items

- *I believe that, by the end of the project, everyone in the team completed roughly the same amount of work.*
- *One or more people in my group did significantly more work than others.****
- *Everyone in the group contributed an amount of work that I consider fair.*[40]
- *Compared to the average amount done by other group members, I did more work.****

Method of Dividing Work

- *All members of my group contributed similar amounts of the various subtasks required by the assignment (writing, design, programming, etc).*

If you wish to clarify, explain, or say anything else about any of your ratings on this page, please use the space below. (Optional)

Free text response box here.

[40]Originally, before pilot: *Everyone in the group contributed a fair amount of work towards the project.*

Page 5—Your satisfaction with your group

The following statements are about your overall satisfaction with your team members, the products of your group, and your satisfaction with the way in which work was organised.

Thinking about your Integrated Project group, to what extent do you agree with each of the following statements?

*(Statements from the following measures were then presented, with order of statements randomised for each participant. Participants gave their response on a seven-point scaled, labelled at the leftmost end with 'Strongly Disagree' and 'Strongly Agree' at the rightmost. Statements with *** were reverse scored.)*

Satisfaction with product:

- *Overall, I am satisfied with the quality of the reports our group produced.*
- *Overall, I am satisfied with the quality of the system our group produced.*

Satisfaction with procedure:

- *I am satisfied with the way in which work was allocated among group members.*
- *Overall, I am satisfied with our chosen method of allocating work to the members of our group.*

Satisfaction with performance:

- *The quality of work produced by the other members of my group was satisfactory to me.*
- *Overall, I am satisfied with the collective performance of my team members.*

Do you have any other comments or anything else you would like to add regarding your satisfaction with the performance of your group members?

You may reference individual group members by name, if you wish. To maintain anonymity in the data, any names you mention will be replaced with pseudonyms during analysis and presentation.

Free text response goes here.

Is there anything else you would like to add about your Integrated Project group that was not covered by the survey? if so, please use the space below to provide details.

For example, if one group member did not contribute any work and had a detrimental impact on your project, you may wish to provide details of that experience here. You may also describe any other experiences or issues that you encountered during the IP unit that may help us to understand your responses.

Again, any names you mention will be replaced with pseudonyms during analysis and presentation.

Free text response goes here.

Page 6—Ending page

Thanks, you've finished!

If you wish to be included in the prize draw, please enter your Bath username so that I will be able to contact you:

Please click the 'Next' button to record your responses.

Thank you very much for completing this survey.

If you have any comments or questions, please drop me an email at: r.m.kelly@bath.ac.uk

Otherwise, good luck in your exams!

Appendix C

Study 2 Materials

Contents: Consent form, briefing scripts, instructions, and questionnaires.

C.1 Consent Form

This consent form was used in all experiments reported in Chapter 5.

Consent Form

Study Overview

This is a study of workload division during collaborative search on the Web. For your participation today you will be paid £5 in cash at the end of the experiment.

During the study you will be working with another person who is located in a separate room elsewhere on campus. You will not be told who the person is either during or after the experiment, and they will not be told who you are either during or after the experiment. The person you will be working with is real and is not the experimenter.

The task you will be working on requires you to complete an information-seeking task involving simple Web searches. During the experiment we will be recording one of your screens using screen-recording software. The software only records information on your screen and it will not be possible for anyone to identify you from this recording.

Important Information

- All data collected during this study will be recorded such that your individual results are anonymous and cannot be traced back to you.
- Your results will not be passed to any third party and are not being collected for commercial reasons.
- Participation in this study does not involve physical or mental risks outside of those encountered in everyday life.
- All procedures and information can be taken at face value and no deception is involved.
- You have the right to withdraw from the study at any time and to have any data about you destroyed. If you do decide to withdraw, please inform the experimenter.
- You will be reimbursed for your time upon successfully completing the study.

By signing this form you acknowledge that you have read the information given above and understand the terms and conditions of this study.

Age:____________________ Gender:____________________

Occupation:____________________ Department: ____________________

Signed: ____________________ Date:____________________

Experimenter: Ryan Kelly, Dept. of Computer Science. RMK22@bath.ac.uk

Supervisor: Professor Stephen J. Payne, Dept. of Computer Science. s.j.payne@bath.ac.uk

C.2 Study 2 Instructions

C.2.1 Proposer Instructions for DLUG

Instructions

In this study you are required to search the Web for information in order to form a reading list for a night class. The reading list must contain 10 reliable sources. Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person will be given the same research topic as you and also has to create a reading list containing 10 sources. You will not be told who the other person is either during or after the study, and they will not be told who you are either during or after the experiment.

The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, you have been given the opportunity to make a decision about how to divide the work. You can do this by specifying how many sources you want the other person partner to look for. You will then be responsible for finding any remaining sources towards the required total of 10.

The way you should do this is as follows:

- The minimum number of sources you can ask the other person to find is 1, and the maximum is 9. You should privately decide how you want to divide the work. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- You will see a chat window on your screen. You should specify the allocation by typing in the following style: *"I want you to find _ sources"*.
- The other person will then have the opportunity to review the allocation you have proposed. They will respond to your allocation with the word *'accept'* or *'reject'*.
- If they accept the allocation, any sources that you both obtain will be shared and your findings will contribute jointly to the required 10 sources.
- If they reject, you will each have to find 10 sources separately.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

C.2.2 Responder Instructions for DLUG

Instructions

In this study you are required to search the Web for information in order to form a reading list for a night class. The reading list must contain 10 reliable sources. Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person will be given the same research topic as you and also has to create a reading list containing 10 sources. You will not be told who the other person is either during or after the study, and they will not be told who you are either during or after the experiment.

The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, the other person has been assigned the task of dividing the workload, and it is up to you to respond to their allocation.

The way you should do this is as follows:

- You will see a workload allocation appear in the chat. This is the number of sources that the other person wants you to find. The minimum number of sources they will ask you to find is 1 and the maximum is 9. The other person will then be responsible for finding the remaining sources for the required total of 10. The allocation will appear in the following style: *"I want you to find _ sources"*.
- If you accept their allocation, any sources that you both find will be shared and your findings will contribute jointly to the required 10 sources.
- If you reject the offer, you will both have to find 10 sources separately.
- You should privately decide whether or not you want to accept their allocation. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- If you are happy with the allocation, please answer: *"Accept"*.
- Or, if you do not accept their allocation, please answer: *"Reject"*.

Please do not type anything else in response to their allocation.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

C.2.3 Web Search Task Instructions

Web Search Task Instructions

The next part of this study requires you to actually create the reading list, in accordance with the outcome of your allocation procedure. The reading list should contain sources that could be used to answer the following question:

> *"To what extent can design be considered a psychological process?"*

To do this you should use the Web to find sources, which can be anything you consider as relevant to the research topic (so books, journals, websites—it is up to you to decide). Each time you find a source you should record it using the chat. You should include a snippet of relevant text from the source material, as well as a hyperlink to the source. The other person will be doing the same during the study.

The only criteria are that:

- Each information source must be unique, so you cannot complete the study by using the same source multiple times.
- You cannot use the same Web page more than once, though you may use multiple pages from the same site (for example, Wikipedia).
- You cannot use items your partner has already shared, but you can search using the same search engines or on the same websites.
- And finally, you can communicate with your partner in any way you wish during this task but you must not reveal any identifying information about yourself.

Once you have completed your own allocation, you will have finished the experiment. Please inform me (the experimenter) when you are finished.

C.3 Study 2 Questionnaires

C.3.1 Proposer Questionnaire

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

Did you understand what you were required to do in the study today? (Yes / No)

Using the space below, please describe what you had to do during the study today. If you did not understand the study, please describe why.

With regard to the allocation you chose to make...

Why did you choose to allocate the work in the way that you did?

Did your partner accept or reject your allocation?

Why do you think they chose to accept / reject?

Did you communicate about the task in order to organise your search activities?

With regard to the question you were asked to research...

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

Do you have any other comments on the study or recommendations for how this study could be improved in the future?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

C.3.2 Responder Questionnaire

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

Did you understand what you were required to do in the study today? (Yes / No)

Using the space below, please describe what you had to do during the study today. If you did not understand the study, please describe why.

With regards to the allocation you were presented with...

Why do you think your partner chose to allocate the work in the way that they did?

Did you accept or reject your partners allocation?

Why did you choose to accept / reject?

Did you communicate about the task in order to organise your search activities?

With regards to the question you were asked to research...

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

Do you have any other comments on the study or recommendations for how this study could be improved in the future?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

APPENDIX D

STUDY 3 MATERIALS

Contents: Briefing scripts, instructions, and questionnaires.

D.1 Study 3 Instructions

D.1.1 Proposer Instructions for DLUG, Pop Music Condition

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about artists or bands that shaped the history of popular music in the 20th century.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person has been given the same research topic as you and also has to create a reading list containing 10 sources. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, you have been given the opportunity to make a decision about how to divide the work. You can do this by specifying how many sources you want the other person partner to look for. You will then be responsible for finding any remaining sources towards the required total of 10.

The way you should do this is as follows:

- The minimum number of sources you can ask the other person to find is 1, and the maximum is 9. You should privately decide how you want to divide the work. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- You will see a chat window on your screen. You should specify the allocation by typing in the following style: *"I want you to find _ sources"*.
- The other person will then have the opportunity to review the allocation you have proposed. They will respond to your allocation with the word *'accept'* or *'reject'*.
- If they accept the allocation, any sources that you both obtain will be shared and your findings will contribute jointly to the required 10 sources.
- If they reject, you will each have to find 10 sources separately.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

D.1.2 Responder Instructions for DLUG, Pop Music Condition

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about artists or bands that shaped the history of popular music in the 20th century.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person has been given the same research topic as you and also has to create a reading list containing 10 sources. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, the other person has been assigned the task of dividing the workload, and it is up to you to respond to their allocation.

The way you should do this is as follows:

- You will see a workload allocation appear in the chat. This is the number of sources that the other person wants you to find. The minimum number of sources they will ask you to find is 1 and the maximum is 9. The other person will then be responsible for finding the remaining sources for the required total of 10. The allocation will appear in the following style: *"I want you to find _ sources"*.
- If you accept their allocation, any sources that you both find will be shared and your findings will contribute jointly to the required 10 sources.
- If you reject the offer, you will both have to find 10 sources separately.
- You should privately decide whether or not you want to accept their allocation. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- If you are happy with the allocation, please answer: *"Accept"*.
- Or, if you do not accept their allocation, please answer: *"Reject"*.

Please do not type anything else in response to their allocation.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

D.1.3 Proposer Instructions for DLUG, Slime Mold Condition

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about the life cycle of mycetozoa slime molds.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person has been given the same research topic as you and also has to create a reading list containing 10 sources. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, you have been given the opportunity to make a decision about how to divide the work. You can do this by specifying how many sources you want the other person partner to look for. You will then be responsible for finding any remaining sources towards the required total of 10.

The way you should do this is as follows:

- The minimum number of sources you can ask the other person to find is 1, and the maximum is 9. You should privately decide how you want to divide the work. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- You will see a chat window on your screen. You should specify the allocation by typing in the following style: *"I want you to find _ sources"*.
- The other person will then have the opportunity to review the allocation you have proposed. They will respond to your allocation with the word *'accept'* or *'reject'*.
- If they accept the allocation, any sources that you both obtain will be shared and your findings will contribute jointly to the required 10 sources.
- If they reject, you will each have to find 10 sources separately.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

D.1.4 Responder Instructions for DLUG, Slime Mold Condition

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about the life cycle of mycetozoa slime molds.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person has been given the same research topic as you and also has to create a reading list containing 10 sources. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, the other person has been assigned the task of dividing the workload, and it is up to you to respond to their allocation.

The way you should do this is as follows:

- You will see a workload allocation appear in the chat. This is the number of sources that the other person wants you to find. The minimum number of sources they will ask you to find is 1 and the maximum is 9. The other person will then be responsible for finding the remaining sources for the required total of 10. The allocation will appear in the following style: *"I want you to find _ sources"*.
- If you accept their allocation, any sources that you both find will be shared and your findings will contribute jointly to the required 10 sources.
- If you reject the offer, you will both have to find 10 sources separately.
- You should privately decide whether or not you want to accept their allocation. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- If you are happy with the allocation, please answer: *"Accept"*.
- Or, if you do not accept their allocation, please answer: *"Reject"*.

Please do not type anything else in response to their allocation.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

D.1.5 Web Search Task Instructions, Both Conditions

Web Search Task Instructions

The next part of this study requires you to actually create the reading list, in accordance with the outcome of your allocation procedure.

To do this you should use the Web to find sources, which can be anything you consider as relevant to the research topic (so books, journals, websites—it is up to you to decide). Each time you find a source you should record it using the chat. You should include a snippet of relevant text from the source material, as well as a hyperlink to the source. The other person will be doing the same during the study.

The only criteria are that:

- Each information source must be unique, so you cannot complete the study by using the same source multiple times.
- You cannot use the same Web page more than once, though you may use multiple pages from the same site (for example, Wikipedia).
- You cannot use items your partner has already shared, but you can search using the same search engines or on the same websites.
- And finally, you can communicate with your partner in any way you wish during this task but you must not reveal any identifying information about yourself.

Once you have completed your own allocation, you will have finished the experiment. Please inform me (the experimenter) when you are finished.

As a reminder, the task is:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about artists or bands that shaped the history of popular music in the 20th century / the life cycle of mycetozoa slime molds.

D.2 Study 3 Questionnaires

D.2.1 Proposer Questionnaire

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

With regard to the allocation you chose to make...

Why did you choose to allocate the work in the way that you did? What factors, if any, influenced your decision?

Did your partner accept or reject your allocation?

Why do you think they chose to accept / reject?

Did you consider making a different offer to the one you eventually proposed? If so, what did you consider?

On reflection, and having now completed the work, would you have made a different offer? Why?

How did you feel about the topic you were asked to research? What was your reaction when you first saw the topic?

With regard to the question you were asked to research...

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

How interesting did you find the topic?

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

How interesting was *the experience of searching for information about the topic?* (nb. This refers to your feelings about the task itself, not the topic you were presented with).

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

Do you have any other comments on the study or recommendations for how this study could be improved in the future?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

D.2.2 Responder Questionnaire

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

With regard to the allocation you chose to make...

Why do you think your partner chose to allocate the work in the way that they did?

Did you accept or reject your partners allocation?

Why did you choose to accept / reject?

What factors, if any, influenced your decision when responding to your partners allocation?

On reflection, and having now completed the work, would you accept the offer they made a second time? Why?

How did you feel about the topic you were asked to research? What was your reaction when you first saw the topic?

With regard to the question you were asked to research...

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

How interesting did you find the topic?

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

How interesting was *the experience of searching for information about the topic?* (nb. This refers to your feelings about the task itself, not the topic you were presented with).

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

Do you have any other comments on the study or recommendations for how this study could be improved in the future?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

Appendix E

Study 4 Materials

Contents: Briefing scripts, instructions, and questionnaires.

E.1 Study 4 Instructions

E.1.1 Dictator Instructions, Complete Information

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about the life cycle of mycetozoa slime molds.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person must also create a reading list on the same topic as you in order to earn their £5. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, you have been given the opportunity to make a decision about how to allocate the work. You can do this by specifying how many sources you want the other person to look for. You will then be responsible for finding any remaining sources towards the required total of 10.

The way you should do this is as follows:

- You must first decide how many sources to allocate to the other person. Any allocation between 0 and 10 is perfectly legitimate for the purposes of this study. This means that you can choose to allocate all 10 sources, none at all, or any number in between. Whatever your allocation, you will be responsible for finding any remaining sources for the required total of 10. Keep in mind that you will not meet the other person and that the allocation does not affect your payment—you will be paid £5 regardless of how much work you leave for yourself and you do not need to wait for them while they work.
- As an example, if you allocate 0 to the other person, you will be required to find 10 by yourself, whereas if you allocate all 10, you will not be required to find any at all and you can leave the study with your money right away. Your task at this point is to decide how much work each person should do to earn his or her £5.
- You should privately decide how you want to divide the work. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- You will see a chat window on your screen. You should specify the allocation by typing in the following style: *"I want you to find _ sources"*.
- Please do not type anything else when making your allocation.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

E.1.2 Dictatee Instructions, Complete Information

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about the life cycle of mycetozoa slime molds.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person must also create a reading list on the same topic as you in order to earn their £5. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, the other person has been assigned the task of allocating the work. They will specify how many sources they would like you to look for, and they will then be responsible for finding any remaining sources towards the required total of 10.

The way this will be done is is as follows:

- You will see a workload allocation appear in the chat—this is the number of sources that the other person has decided to allocate to you. This number will be between 0 and 10.
- The allocation will appear in the following style:
- *"I want you to find _ sources"*.

This will be the amount of work you are required to do to get your £5 payment. Note that they may assign all of the workload to you, in which case they will not respond via the chat when you begin your work on the task.
Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

E.1.3 Dictator Instructions, Incomplete Information

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about the life cycle of mycetozoa slime molds.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person must also create a reading list on the same topic as you in order to earn their £5. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, you have been given the opportunity to make a decision about how to allocate the work. You can do this by specifying how many sources you want the other person to look for. You will then be responsible for finding any remaining sources towards the required total of 10.

The way you should do this is as follows:

- You must first decide how many sources to allocate to the other person. Any allocation between 0 and 10 is perfectly legitimate for the purposes of this study. This means that you can choose to allocate all 10 sources, none at all, or any number in between. Whatever your allocation, you will be responsible for finding any remaining sources for the required total of 10. Keep in mind that you will not meet the other person and that the allocation does not affect your payment—you will be paid £5 regardless of how much work you leave for yourself and you do not need to wait for them while they work.
- As an example, if you allocate 0 to the other person, you will be required to find 10 by yourself, whereas if you allocate all 10, you will not be required to find any at all and you can leave the study with your money right away. Your task at this point is to decide how much work each person should do to earn his or her £5.
- You should privately decide how you want to divide the work. Please keep the decision to yourself and do not inform me (the experimenter) of your decision.
- You will see a chat window on your screen. You should specify the allocation by typing in the following style: *"I want you to find _ sources"*.
- Please do not type anything else when making your allocation.

Before making your allocation, you should also be aware that the other person does not know the total number of sources required for the reading list. Instead, we have told them that they must find some number of sources and that you will be allocating the work to them. This means that when you make the allocation, they will know their own assignment but will have no way of being certain about yours.

Once you have made your allocation, please inform me (the experimenter) and you will be introduced to the second part of the study.

E.1.4 Dictatee Instructions, Incomplete Information

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have reliable sources containing information about the life cycle of mycetozoa slime molds.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person must also create a reading list on the same topic as you in order to earn their £5. The first part of this study involves making a proposal about how to divide the workload between the two of you. On this occasion, the other person has been assigned the task of allocating the work. They will specify how many sources they would like you to look for, and they will then be responsible for finding any remaining sources towards the required total.

The way this will be done is is as follows:

- You will see a workload allocation appear in the chat—this is the number of sources that the other person has decided to allocate to you. This number will be between 0 and 10.
- The allocation will appear in the following style:
- *"I want you to find _ sources"*.

This will be the amount of work you are required to do to get your £5 payment. Note that they may assign all of the workload to you, in which case they will not respond via the chat when you begin your work on the task.

Once you have done this, please inform me (the experimenter) and you will be introduced to the second part of the study.

E.1.5 Web Search Task Instructions, Both Conditions

Web Search Task Instructions

The next part of this study requires you to actually create the reading list, in accordance with the outcome of your allocation procedure.

To do this you should use the Web to find sources, which can be anything you consider as relevant to the research topic (so books, journals, websites—it is up to you to decide). Each time you find a source you should record it using the chat. You should include a snippet of relevant text from the source material, as well as a hyperlink to the source. The other person will be doing the same during the study.

The only criteria are that:

- Each information source must be unique, so you cannot complete the study by using the same source multiple times.
- You cannot use the same Web page more than once, though you may use multiple pages from the same site (for example, Wikipedia).
- You cannot use items your partner has already shared, but you can search using the same search engines or on the same websites.
- And finally, you can communicate with your partner in any way you wish during this task but you must not reveal any identifying information about yourself.

Once you have completed your own allocation, you will have finished the experiment. Please inform me (the experimenter) when you are finished.

As a reminder, the task is:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about the life cycle of mycetozoa slime molds.

[In the Incomplete Information condition, the passage given to responders read: *You have been asked to form a reading list for students of a night class. The reading list must have reliable sources containing information about the life cycle of mycetozoa slime molds.*]

E.2 Study 4 Questionnaires

E.2.1 Dictator Questionnaire, Complete Information

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

With regard to the way you went about completing this study:

Why did you choose to allocate the work in the way that you did? What factors, if any, influenced your decision?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied were you with the allocation of work?

On reflection, and having now completed the task, would you have organised the work differently? If so, why?

Did you consider making a different offer to the one you eventually proposed? If so, what did you consider?

If you did not do any work on the search task then please skip the next three questions.

Did the task you were asked to complete remind you of any real-world tasks that you have done to find

information in collaboration with other people?

Did you pay attention to the activities of your partner during the study? If so, what did you do?

Did you feel influenced by the activities of your partner during the study? If so, how?

Regarding the topic of mycetozoa slime mold:

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

If you worked alongside the other person on this topic:

How interesting did you find the topic?

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

How interesting was *the experience of searching for information about the topic?* (nb. This refers to your feelings about the task itself, not the topic you were presented with).

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

If you worked alongside the other person on the topic:

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the quality of the sources produced by your partner during the study?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the speed at which your partner worked during the study?

Was their speed too fast, too slow, or about right?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

E.2.2 Dictatee Questionnaire, Complete Information

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

With regard to the way you went about completing this study:

Why do you think the other person chose to allocate the work in the way that they did?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied were you with the allocation of work?

Imagine that this study had been as follows: the other person makes you a workload offer, and you either accept or reject it. If you accept, you each do the work as assigned. If you reject, you each obtain 10 items separately. Given the allocation that the other person made to you today, would you accept or reject it in such a setting? Why?

If you had been responsible for dividing the work in this study, what would your allocation have been?

If you did not do any work on the search task then please skip the next three questions.

Did the task you were asked to complete remind you of any real-world tasks that you have done to find

information in collaboration with other people?

Did you pay attention to the activities of your partner during the study? If so, what did you do?

Did you feel influenced by the activities of your partner during the study? If so, how?

Regarding the topic of mycetozoa slime mold:

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

If you worked alongside the other person on this topic:

How interesting did you find the topic?

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

How interesting was *the experience of searching for information about the topic?* (nb. This refers to your feelings about the task itself, not the topic you were presented with).

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

If you worked alongside the other person on the topic:

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the quality of the sources produced by your partner during the study?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the speed at which your partner worked during the study?

Was their speed too fast, too slow, or about right?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

E.2.3 Dictator Questionnaire, Incomplete Information

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

With regard to the way you went about completing this study:

Why did you choose to allocate the work in the way that you did? What factors, if any, influenced your decision?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied were you with the allocation of work?

On reflection, and having now completed the task, would you have organised the work differently? If so, why?

Did the fact that the other person did not know the amount of work required affect your allocation? Why?

Did you consider making a different offer to the one you eventually proposed? If so, what did you consider?

If you did not do any work on the search task then please skip the next three questions.

Did the task you were asked to complete remind you of any real-world tasks that you have done to find

information in collaboration with other people?

Did you pay attention to the activities of your partner during the study? If so, what did you do?

Did you feel influenced by the activities of your partner during the study? If so, how?

Regarding the topic of mycetozoa slime mold:

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

If you worked alongside the other person on this topic:

How interesting did you find the topic?

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

How interesting was *the experience of searching for information about the topic?* (nb. This refers to your feelings about the task itself, not the topic you were presented with).

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

If you worked alongside the other person on the topic:

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the quality of the sources produced by your partner during the study?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the speed at which your partner worked during the study?

Was their speed too fast, too slow, or about right?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

E.2.4 Dictatee Questionnaire, Incomplete Information

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

With regards to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

With regard to the way you went about completing this study:

Why do you think the other person chose to allocate the work in the way that they did?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied were you with the allocation of work?

Imagine that this study had been as follows: the other person makes you a workload offer, and you either accept or reject it. If you accept, you each do the work as assigned. If you reject, you each obtain 10 items separately. Given the allocation that the other person made to you today, would you accept or reject it in such a setting? Why?

If you had been responsible for dividing the work in this study, what would your allocation have been?

If you did not do any work on the search task then please skip the next three questions.

Did the task you were asked to complete remind you of any real-world tasks that you have done to find

information in collaboration with other people?

Did you pay attention to the activities of your partner during the study? If so, what did you do?

Did you feel influenced by the activities of your partner during the study? If so, how?

Regarding the topic of mycetozoa slime mold:

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

If you worked alongside the other person on this topic:

How interesting did you find the topic?

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

How interesting was *the experience of searching for information about the topic?* (nb. This refers to your feelings about the task itself, not the topic you were presented with).

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

If you worked alongside the other person on the topic:

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the quality of the sources produced by your partner during the study?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the speed at which your partner worked during the study?

Was their speed too fast, too slow, or about right?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

APPENDIX F

STUDY 5 MATERIALS

Contents: Briefing scripts, instructions, and questionnaires.

F.1 Study 5 Instructions

F.1.1 Pop Music Topic

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about artists or bands that shaped the history of popular music in the 20th century.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person has been given the same research topic as you and also has to create a reading list containing 10 sources.

You must use the Web to find sources for the reading list. Sources can be anything you consider as relevant to the research topic (so books, journals, websites—it is up to you to decide). Each time you find a source you should record it using the chat. You should include a snippet of relevant text from the source material, as well as a hyperlink to the source. The other person will be doing the same during the study.

The only criteria are that:

- Each information source must be unique, so you cannot complete the study by using the same source multiple times.
- You cannot use the same Web page more than once, though you may use multiple pages from the same site (for example, Wikipedia).
- You cannot use items your partner has already shared, but you can search using the same search engines or on the same websites.
- And finally, you can communicate with your partner in any way you wish during this task but you must not reveal any identifying information about yourself.

Once you have completed your own part of the task, you will have finished the experiment. Please inform the experimenter when you are finished.

F.1.2 Slime Mold Topic

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about the life cycle of mycetozoa slime molds.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person has been given the same research topic as you and also has to create a reading list containing 10 sources.

You must use the Web to find sources for the reading list. Sources can be anything you consider as relevant to the research topic (so books, journals, websites—it is up to you to decide). Each time you find a source you should record it using the chat. You should include a snippet of relevant text from the source material, as well as a hyperlink to the source. The other person will be doing the same during the study.

The only criteria are that:

- Each information source must be unique, so you cannot complete the study by using the same source multiple times.
- You cannot use the same Web page more than once, though you may use multiple pages from the same site (for example, Wikipedia).
- You cannot use items your partner has already shared, but you can search using the same search engines or on the same websites.
- And finally, you can communicate with your partner in any way you wish during this task but you must not reveal any identifying information about yourself.

Once you have completed your own part of the task, you will have finished the experiment. Please inform the experimenter when you are finished.

F.1.3 Art Crime Topic

Instructions

In this study you are required to search the Web for information in order to complete the following task:

You have been asked to form a reading list for students of a night class. The reading list must have 10 reliable sources containing information about subject of international art crime, focusing on instances of fraud or theft in the international buying and selling of art.

Rather than complete this task alone, you have the opportunity to work with another person who is located elsewhere on campus. This other person has been given the same research topic as you and also has to create a reading list containing 10 sources.

You must use the Web to find sources for the reading list. Sources can be anything you consider as relevant to the research topic (so books, journals, websites—it is up to you to decide). Each time you find a source you should record it using the chat. You should include a snippet of relevant text from the source material, as well as a hyperlink to the source. The other person will be doing the same during the study.

The only criteria are that:

- Each information source must be unique, so you cannot complete the study by using the same source multiple times.
- You cannot use the same Web page more than once, though you may use multiple pages from the same site (for example, Wikipedia).
- You cannot use items your partner has already shared, but you can search using the same search engines or on the same websites.
- And finally, you can communicate with your partner in any way you wish during this task but you must not reveal any identifying information about yourself.

Once you have completed your own part of the task, you will have finished the experiment. Please inform the experimenter when you are finished.

F.2 Study 5 Questionnaire

This questionnaire was spread over three pages and participants had ample room for free-text responses.

Questionnaire

With regard to your computer expertise...

How would you rate your computer expertise?

Novice, Intermediate, Expert

On average, how often do you use the Web to search for information?

Once every couple of months, A few times per week, Once per day, Several times per day

With regard to the way you went about completing this study:

On reflection, and having now completed the task, would you have organised the work differently? Why?

What factors, if any, did you consider when deciding how to organise the work?

Did you monitor or pay attention to the activities of your partner during the study? If so, what did you do?

Did you feel influenced by the activities of your partner during the study? If so, how?

To what extent did you feel pressured by the activities of your partner during the study? For example, did you feel the need to speed up or slow down your own work rate in order to match their speed?

How did you decide whether or not a source should be included in the reading list?

Was your decision to include sources affected by the speed at which you and your partner were working?

If so, how?

Regarding the topic you were asked to research:

How familiar are you with the subject matter of the topic you were asked to research?

Very unfamiliar, Slightly unfamiliar, About average, Slightly familiar,Very familiar

How difficult was it to find sources relevant to this topic?

Very easy, Slightly easy, About average, Slightly difficult, Very difficult

How interesting did you find the topic?

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

How interesting was *the experience of searching for information about the topic?* (nb. This refers to your feelings about the task itself, not the topic you were presented with).

Very dull, Slightly dull, About average, Slightly interesting, Very difficult

If you worked alongside the other person on the topic:

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the quality of the sources produced by your partner during the study?

On a scale of 1 to 7, where 1 is highly unsatisfied and 7 is highly satisfied, how satisfied are you with the speed at which your partner worked during the study?

Was their speed too fast, too slow, or about right?

Thank you for taking the time to complete this study. If you have any questions, please feel free to ask the experimenter.

Appendix G

Study 6 Materials

Contents: Briefing script and semi-structured interview protocol.

G.1 Interview Briefing Script

The following script was read to each participant prior to the commencement of each interview.

Thank you for participating in our study, today I will be asking you questions based on your use of Coagmento/Diigo. I'll be asking questions about a range of issues, including the process you followed during your searches, how you communicated with your partner, and I'm also interested in your opinions of the tool we gave you. Before beginning, I would like to tell you a few things about the interview — nothing bad, just some information to put you at ease.

- First of all, please feel free to be frank and honest when talking about your experiences with the system. We didn't build the tool and we have no reason to be offended by anything you say about it, so please feel free to speak truthfully and be honest about your experiences. Your opinions are valuable to us.
- Also, there are no right or wrong answers in this study. So please feel free to share your opinions and ideas, even if you think they might be silly. They probably aren't.
- I have basic framework of 32 questions but sometimes they get answered during conversation, so I may not need to ask all of them. The interview shouldn't last more than an hour in total. We can follow up on your experiences if there is something particular that you want to talk more about.
- Is it okay if we record the interview? [Await confirmation]. We will not share the recording with anyone and when your responses are transcribed it will not be possible for anyone to identify you from the transcript.

Do you have any questions before we begin?

G.2 Interview Questions

The list of questions here was informed by the literature on CIS reviewed in Chapter 6; the list of questions used by Capra *et al*. (2010); and the salient concerns of this thesis. (Fairness, division of labour, awareness, etc.) Please note also that these are guideline questions; invariably, some would be answered during the course of talking about other questions, and were therefore not asked.

Overview of the Project

I'd like to start by asking you some general questions about your search activity during the study.

1. What type of information were you and your partner looking for?
2. What was the overall goal of the project? Why were you searching?
3. Where did most of the searches take place?
4. Did you search at the same time as your partner or at different times? Were you together or apart?
5. How long did it take to complete the task?
6. How many search sessions were there? (Ask for rough guess if the participant struggles, then verify the guess by cross-checking with search logs and partner's statements).

7. How would you normally go about looking for information when doing tasks like the one you did during the study? If you weren't participating in this study, what would you do? (Try to return to this question when participant talks about behaviours during the interview).

Information Seeking

Let's talk about the process you followed when searching for information.

8. Can you give me a general idea of your approach towards achieving your goal? For example, how did you get started on the task?
9. How did you organize the task? Was there any sort of division of labour at all?
10. What challenges, if any, did you encounter related to searching and managing results found?
11. What did you do with information once you had found it?
12. Did you encounter any problems during the execution of the task?
13. How did you resume your searches from previous sessions? How did you remember where you had left off?

Communication

Let's talk about how and your partner communicated and exchanged information regarding your chosen task.

14. How did you share information with your partner?
15. How did they share information with you?
16. How did you and your partner communicate during the time you were working on your tasks? For example, what methods did you use to discuss what you had found?
17. Did you ever communicate *during* searches?
18. How easy was it for you to understand what your partner had already looked at during their work?

Tool Use

Let's talk about the tool.

19. Thinking back to your use of the toolbar, which features did you use most frequently? (Ask them to elaborate on what they used them for. Ask them about each bookmarking feature and whether or not they used them).
20. Do you remember choosing not to use any particular features? (If so, why didnt they use them?)
21. Did you use any of the toolbars communication features, for example the chat or sidebar?
22. Did you encounter any problems during your use of the system? (If so, what were they?)

23. Were there any ways in which the tool did not support your activities?
24. Was there anything you felt the tool did badly?
25. Did you use any other tools or methods to capture information? (e.g. paper notes)
26. Can you give me an opinion of what you thought about the toolbar, overall?

Ending the Process

27. Would you say that your workloads during this task were roughly equal?
28. Was fairness important to you? Were you concerned about the amount of work completed by each person?
29. How did you decide that the quantity of information you found was enough? (As in, at what point did they decide to terminate information seeking activities?)
30. How did you achieve consensus regarding the outcome of the work? As in, how did you decide which option to choose? (Which hotel or destination to select)
31. Overall, how would you describe your success in achieving what you wanted to achieve by using the tool?
32. Is there anything else you would like to add that we have not covered?

Appendix H

Study 7 Materials

Contents: List of web streams observed, briefing script, and semi structured interview protocol.

H.1 List of Web Streams

These streams were observed by the author of this thesis to learn more about raiding in Warcraft at the time of study (August 2013). Note that consent was not obtained to observe these players as they are broadcasting via public channels, i.e. they invite and encourage observation of their activity.
Channel for observation of groups doing player versus player battlegrounds.
www.twitch.tv/braindeadly

Channels for observations of top-tier raiding activity in Throne of Thunder and other raiding zones.
www.twitch.tv/fragnance
www.twitch.tv/treckie
www.twitch.tv/june_tv
www.twitch.tv/estarra

H.2 Interview Briefing Script

The following script was read to each participant prior to the commencement of each interview.

Thank you for participating in our study. I will be asking you questions based on your experiences of playing World of Warcraft. I'll be asking questions about a range of issues, including the roles you assume while raiding and your responsibilities in your group, how you stay aware of team members' contributions, and the issues that arise through use of various interface metrics. Before beginning, I would like to tell you a few things about the interview—nothing bad, just some information to put you at ease.

- First of all, please feel free to be frank and honest when talking about your experiences. There are no right or wrong answers in this study, so please feel free to share your opinions and ideas, even if you think they might not be relevant.
- I have a basic framework of 29 questions but sometimes they get answered during conversation, so I may not need to ask all of them. The interview shouldn't last more than an hour in total. We can follow up on your experiences if there is something particular that you want to talk more about.
- Is it okay if we record the interview? [Await confirmation]. We will not share the recording with anyone and when your responses are transcribed it will not be possible for anyone to identify you from the transcript.

Do you have any questions before we begin?

H.3 Interview Questions

About your Warcraft Character

I'd like to start by asking you some general questions about you and your experiences with WoW.

1. What is your age?
2. What is your gender?
3. Which country is your current permanent place of residence?

4. For how long have you been playing Warcraft? (Years/months please estimate).
5. What is the race of the character you play most on Warcraft? (i.e. your primary character that you use for raiding purposes).
6. What is the class of the character you play most on Warcraft? (again the character you use for raiding purposes).
7. Are you currently in a guild?
8. How big is the guild?
9. How would you describe the guild? (e.g. casual raiding guild, medium sized raiding guild, hardcore raiding guild).
10. Are you in a position of authority within that guild? (E.g. guild officer).
11. How frequently do you participate in raids with members of that guild?
12. What is your functional role while raiding? (Damage dealer, healer, tank, other - If other, how would you describe the role?)
13. Do you have an organizational role while raiding? (E.g., raid leader, communications officer).
14. What are the demands of this role? What does it require you to do while raiding? What are your chief responsibilities?

General Awareness Issues

Let's talk about more specific aspects of raiding now.

15. While you are on a raid, what are the primary duties for you in your class role?
16. While you are on a raid, to what extent is it necessary for you to pay attention to what your team mates are doing? Why?
17. How do you stay aware of what your team mates are doing? What sort of information are you looking for / monitoring?
18. How does knowing what they are doing help you on a raid?

Interface

19. What interface elements do you use to support your raiding activities, if any?
20. How do these elements help you? (Specifically what aspects of raiding do they support?)
21. Do you use damage / healing / threat meters?
22. How do these meters help you during a raid?
23. Have you encountered any issues surrounding use of these meters? E.g. positive or negative side effects?

Participation and Contributions from Team

24. When you're raiding, how do you decide how much effort should be invested at any given time?

25. How do you ensure that everyone is participating in a raid?

26. How does your raid group or guild decide how drops and loot should be assigned? How do you decide who gets which item?

27. Do the contents of the damage or healing meters help you to determine who should get which loot drops?

28. Is it important that all team members contribute equally? How do you know whether someone is making an appropriate contribution to the raid?

29. Is there anything else you'd like to add?

End Interview.

Appendix I

An Example Raid Encounter

Contents: Description of the Onyxia raid encounter from World of Warcraft.

Defeating Onyxia: An Example Raid Encounter

Since the reader may not be familiar with the nitty-gritty of raiding in Warcraft, we offer a short account of how a typical boss encounter plays out. We describe an encounter with the dragon Onyxia, as pictured in Figure I.1. This helps to illustrate the coordination challenges experienced during raid encounters. The Onyxia encounter is from the classic version of Warcraft, and plays out as follows:

- Phase One: This involves a simple 'tank and spank' task where the group's primary tank engages the boss and the rest of the raid deals damage to lower the dragon's health. Because the dragon has some abilities that can potentially harm the raid—for example, she can breath conical fire and can apply cleaved damage with her claws and tail—two very simple coordinated behaviours are required. First, the main tank draws the dragon towards a wall to prevent her from hitting other raid members. Second, DPS classes will attack the dragon from its sides, rather than behind; this minimises the chance of dying from the dragon's tail swipe.

- Phase Two: When Onyxia's health is lowered to 65%, she takes off and becomes airborne in the centre of the room (as depicted in Fig. I.1). At this point, she begins shooting fireballs at random members of the raid, and smaller dragon whelps begin to spawn from either side of the cave. The raid must therefore divert much of their attention away from the main boss and kill the whelps before the room becomes overwhelmed. This task is typically dedicated to tanks and melee-based DPS characters. The remainder of the raid then continues to inflict damage to Onyxia using ranged attacks.

- Phase Three: When Onyxia's health reaches 40%, she flies to the ground and adopts the same pattern of behaviour as in Phase One. However, the raid is faced with two additional challenges: Onyxia bellows an intermittent roar, which causes all members of the raid to be 'feared' for up to three seconds (a feared character runs around randomly and cannot be controlled by its user until the fear wears off or is removed). The raid must overcome this by removing the main tank's fear status as soon as possible to ensure that Onyxia does not wander around and begin killing other raid members. A second challenge is that, when Onyxia roars, eruptions of flame appear through cracks in the floor. Thus raid members are susceptible to increased damage while feared, placing extra pressure on the raid's healers.

Figure I.1: A Warcraft raid in progress. A large group of players is attempting to slay the dragon Onyxia.

In addition to these three phases, Onyxia has a special 'deep breath' ability that can inflict massive damage to the raid. Onyxia is rumoured to be more likely to perform this ability if the raid members clump together; thus, players must try to ensure equidistant spacing throughout each of the aforementioned phases. If the raid is successful in navigating these phases, the raid should defeat Onyxia and will be able to obtain loot from her corpse. Onyxia is actually one of the more simple raid encounters, harder raids may present more difficult coordination challenges by, for example, requiring players to position their avatars in specific places, change positions during the raid, switch from damage to healing, or to perform actions in prescribed sequences.

Once Onyxia's health points reach zero, she dies and the raid is successful. Players can loot her corpse to obtain their rewards.

APPENDIX J

PHOTOGRAPH AND IMAGE CREDITS

Contents: List of necessary accreditations and attributions for images used in this thesis.

J.1 Image Credits

Chapter 6

Figure 6.1, Coagmento is a copyright (c) of Chirag Shah, 2012.
http://www.coagmento.org

Figure 6.2, Diigo is a copyright (c) of Diigo Inc., 2012.
http://www.diigo.com

Chapter 7

Figure 7.1
From Erickson, Halverson, Kellogg, Laff, & Wolf (2002), copyright (c) Association for Computing Machinery.

Other

Figure I.1, "Onyxia raid" copyright (c) by aigarius on Flickr.
http://www.flickr.com/photos/aigarius/4351914262